ANALÍTICA

Revista de Filosofía
Analítica (5), oct. 2025 - sept. 2026
ISSN-L 2805-1815

ANALÍTICA

No. 5, oct. 2025 - sept. 2026, ISSN-L 2805-1815

Revista de Filosofía

Publicación anual

Analítica 5

Revista de Filosofía

Publicación anual

Departamento de Filosofía

Universidad de Panamá

Octubre 2025 – septiembre 2026

ISSN – L 2805 – 1815

Portada: E. López, (sin título), acuarela:

Analítica es una publicación anual en línea del Departamento de Filosofía de la Universidad de Panamá. Nuestra política editorial e instrucciones para los autores se encuentran disponibles en https://revistas.up.ac.pa/index.php/analitica/autores .

Información de contacto: Departamento de Filosofía, Campus Octavio Méndez Pereira, Universidad de Panamá. Tel. +507-5236602. Correo electrónico: revista.analiticafh@up.ac.pa

Autoridades de la Universidad de Panamá

Dr. Eduardo Flores Castro
Rector

Dr. José Emilio Moreno
Vicerrector Académico

Dr. Jaime Javier Gutiérrez
Vicerrector de Investigación y Postgrado

Mgter. Arnold Muñoz
Vicerrector Administrativo

Mgter. Mayanín Rodríguez
Vicerrectora de Asuntos Estudiantiles

Mgter. Ricardo Him Chi
Vicerrector de Extensión

Mgter. José Luis Solís
Director General de Centros Regionales Universitarios y Extensiones Docentes

Mgter. Ricardo A. Parker
Secretario General

Dr. Olmedo García Chavarría
Decano, Facultad de Humanidades

Dr. Franz Poveda
Director, Departamento de Filosofía

Índice

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815

Presentación

En este quinto número, Analítica ofrece a la comunidad filosófica nacional e internacional diecisiete artículos sobre temas de interés tanto para especialistas como para el gran público. Los autores proceden de diversas latitudes del planeta, lo cual es un hecho que no puede soslayarse: muestra que –pese a la visión utilitarista arraigada en el presente, y que no duda en afirmar que la filosofía no sirve para nada– hay quienes desde distintas coordenadas culturales y tradiciones de pensamiento se esmeran de manera decidida por indagar críticamente sobre diversos problemas, teóricos y prácticos, que – simplemente – no pueden ser eludidos.

En la introducción de la *Crítica del juicio,* señala Kant que la filosofía se puede dividir en *teórica* y *práctica*. Desde luego, no se trata de postular una oposición infranqueable entre una y otra, todo lo contrario: lo teórico y lo práctico están íntimamente relacionados, son dos caras de una sola moneda, la de la razón. Podría decirse que esta distinción aplica de manera inequívoca al contenido de este número de Analítica.

Ciertamente, desde una perspectiva teórica se abordan cuestiones como el escepticismo (*Meta-philosophical Skepticism, Self-defeat and Pragmatic Justification,* Shih-Hao Liu), la paradoja del mentiroso (*Semantics of the Liar Paradox,* C. P. Hertogh), contextualismo y composicionalidad (*Radical Contextualism and Open Compositionality: a Metatheoretical Proposal*, Pablo David Chávez Carvajal), causalidad (*Is Motor Representation a Potential Answer to the Problem of Causal Deviance?,* Yukun Chen), la relación entre el lenguaje y lo místico (*El lenguaje y lo místico en el primer Wittgenstein*, Javier Antonio Torres-Vindas), historia de la filosofía (*Comentarios sobre la crítica de Schopenhauer a Kant con relación a la cosa en sí y la ley de causalidad*, Ruling Barragán).

Desde una perspectiva práctica, se abordan tópicos como la responsabilidad (*Moral Responsability in Plato's Philosophy. Commentary on "Timaeus" 86d-87b*, Marc Zapata Pedrosa), la ética aplicada (*Marcos éticos en la ingeniería civil*, Gabriel Montúfar), el pensamiento crítico (*Pensamiento crítico: una visión desde la filosofía*, José Mathurín), la relación entre ética y moral (*Distinción y complementariedad entre ética y moral. Un enfoque desde la filosofía moral*, Alessandro Caviglia), la identidad de género desde la perspectiva feminista (*From Essence to Construction: Feminist Revisions of Identity*, Aiswarya Pradeep Kumar), el sufrimiento (*Reframing Suffering: Buddhist Mindfulness Techniques as Tools for Philosophical Counselling*, Richa Singh), filosofía latinoamericana (*Williams and Dussel on Opacity: Toward a Non-totalizing Method*, Chris Sawyer), conservadurismo (*Fundamentos del conservadurismo scrutoniano*, Joshua Isaac Ramírez Donner).

En otros tópicos, lo teórico y lo práctico se solapan; así, tenemos: epistemología virtuosa y educación (*Virtue Epistemology and Education*, Rashad Rehman y Hassan Ahmad), renta básica e inteligencia artificial (*Basically Intelligent: Ontological and Rationalistic Perspectives on Artificial Intelligence in the Context of Basic Income Discourse*, Shawn Christopher Vigil), antropología filosófica y conocimiento (*Ser humano y conocimiento en las obras tempranas de George Siemens*, Freddy Varona Domínguez)

Como se ha señalado, cada texto es expresión de distintas tradiciones filosóficas. Pese a las diferencias de enfoques, hay un denominador común: la rigurosidad en el tratamiento de los temas y en el desarrollo de los argumentos. Por esta razón, consideramos que el presente número de Analítica ofrece a los lectores e investigadores de filosofía un valioso arsenal conceptual para la reflexión de apremiantes cuestiones que aquejan a las sociedades actuales.

Francisco Díaz Montilla, PhD
Director/Editor Jefe

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 9-27

Artículo/Article

Virtue epistemology and education

Epistemología virtuosa y educación

Rashad Rehman

Franciscan University of Steubenville, United States
rrehman@franciscan.edu
https://orcid.org/0000-0002-7587-4170

Hassan Ahmad

University of British Columbia, British Columbia, Canada
ahmad@allard.ubc.ca
https://orcid.org/0000-0001-6114-5885

DOI https://doi.org/10.48204/2805-1815.8470

INFORMACIÓN DEL ARTÍCULO

Recibido el: 16/5/2025

Aceptado el: 20/8/2025

Keywords:

Education, capability, welfare, intellect, knowledge

Palabras clave:

Educación, capacidad, bienestar, intelecto, conocimiento

ABSTRACT/RESUMEN

Abstract:

Education is a bifurcated process by which knowledge is generated and transferred through learning tangible skills and intangible virtues. This paper uses Amartya Sen's capability approach to advocate for a normative quality of welfare pursuant to the capability of education. For normative welfare, a particular state of consciousness or *midfare* must first be achieved. We characterize this *midfare* as *Intellect*, which is defined by a set of non-exhaustive virtues that we devise and that prioritize the Finnisian value of knowledge over all other subjective pursuits. We also assess a possible relativist critique to *Intellect*. The desire is that a working model of *Intellect* will be implemented within transitioning curricula.

Resumen:

La educación es un proceso bifurcado a través del cual se genera y transfiere conocimiento mediante el aprendizaje de habilidades tangibles y virtudes intangibles. Este artículo utiliza el enfoque de las capacidades de Amartya Sen para abogar por una calidad normativa del bienestar, en consonancia con la capacidad de la educación. Para lograr un bienestar normativo, primero se debe alcanzar un estado de conciencia o punto intermedio. Caracterizamos este punto intermedio como Intelecto, definido por un conjunto de virtudes no exhaustivas que ideamos y que priorizan el valor finnisiano del conocimiento sobre todas las demás búsquedas subjetivas. También evaluamos una posible crítica relativista al Intelecto. El objetivo es que se implemente un modelo funcional del Intelecto en los currículos en transición

Introduction

In recent years, as the process by which knowledge is generated, kept, and transferred within and between individuals and groups, education has attracted a significant amount of scholarship. (See Davidson, 1990; Graves, 1999, 1911; Parker, 1970; Ulich, 1945; Watkins, 2012). In inculcating knowledge, education is the process for achieving truth. Acquiring knowledge through education is a bifurcated process. It has the ability to teach tangible skills such as literacy, numeracy, observatory skills, understanding, critical analysis and reflection. It can also instil intangible virtues including, but not limited to, tolerance, respect, empathy, dignity, and temperance. Other virtues can include courage, moderation, justice, generosity, expansive hospitality, greatness of soul, mildness of temper, truthfulness, easy grace, proper judgment, and practical wisdom, among others. (For a fuller discussion on Aristotelian virtues, see Nussbaum, 1993 and Aristotle, 1998; for lesser discussed virtues e.g., the virtues of conviviality, hospitality, lightheartedness, warmheartedness, et cetera, see DeMarco, 2000).

At times, this bifurcated process runs in parallel tracks. Occasionally, it runs orthogonal. By a 'parallel' bifurcated process we mean that skills and virtues are taught independently of each other. In orthogonal bifurcated processes, skills and values are interrelated such that one affects the other and vice versa. In the discussion that follows, we take for granted that the predominant realm of teaching tangible skills is the domain of formal educational institutions – though it need not be.[1] Our interest is to elaborate upon the orthogonal track to advocate for a particular set of virtues that informs skills taught in formal institutions. We do not specify any target group for our analysis but recognize it can be implemented earliest at the primary level and be most effective at the middle and secondary levels of schooling. During all those periods, students are dependent on a teacher or parent for inculcating skills and virtues in order to perceive the physical and metaphysical world. By 'metaphysical,' we simply mean (descriptively) the comprehensive, systematic and unificatory attempt to develop a worldview that inculcates (possibly) more than the physical world. In other words, we might say, instead of 'metaphysical,' 'the possibilities afforded by the pursuit of knowledge.'

The notion of education as being a public good that is, in part, regulated by the State has historically attracted criticism across the spectrum of political philosophers. While stressing that education is an important socialization mechanism for raising class consciousness, Weberian and Marxist approaches have critiqued State-run education systems for reproducing inherently unequal social structures and reinforcing the ruling class's hegemony (Halvorsen, 1990). On the other end of the political spectrum, John Stuart Mill repudiated the notion of an education system entirely run by the State. He mentioned in his treatise *On Liberty* (Mill, 1991, p. 68), "an education established and controlled by the State should only exist, if it exists at all, as one among many competing experiments, carried on for the purpose of example and stimulus to keep the others up to a certain standard of excellence". (Unlike Mill, our notion of 'excellence' will be rooted in natural law rather than utilitarian theoretical commitments). Despite (some) legitimate worries or criticisms of the State's role in education[2], the overwhelming practice after the French and American revolutions was that education serves a public function and should be promulgated, as least in part, by the State. For the herein purposes, we acknowledge, *descriptively*, that the State plays a role in educating its citizens.[3]

The analysis presented here uses Amartya Sen's theory (1999) that welfare is correctly characterized through a capability lens.[4] This refers to the range of functionalities[5] or ability to achieve a certain goal (Sen, 1999; Cohen, 1993). We apply Sen's theory to the right to education and argue that education as a means to welfare, as Sen conceives it, is incomplete. (We rely here on an account of rights theory situated within the natural law tradition. See e.g., Oderberg, 2013, pp. 375-386). An intermediary step is present and, in fact, needed. G. A. Cohen names that step as *midfare*. In Cohen's characterization, *midfare* lies between capability and welfare (or utility). It is constituted by the mental states and virtues created within the person in order to result in a best level of utility. Therefore, *midfare* is a normative concept that focuses on a capability's qualitative ability to effect varying levels of utility (Cohen, 1993). For education, the specific *midfare* that ought to be imparted from teacher to student is what we term *Intellect*.[6] We choose the term *Intellect* from the Aristotelian conception in the

Nichomachean Ethics (Aristotle, 1998), which deems intellectual virtues such as perceptiveness to be normatively correct responses (Nussbaum, 1993). This means that *Intellect* is a *method* because it is an approach from which one can engage in education, since (we argue) education *assumes* the significance of intellectual virtue.

We stipulatively define *Intellect* as the particular set of virtues learned and continually expanded upon through an education absent of prejudice, epistemic bias, and unwarranted beliefs. It is based on generally accepted observable facts, evidence, and experiential realities.[7] In line with the Finnisian conception, the set of virtues normatively learned and inculcated as part of *Intellect* are done in pursuit of the overarching basic value of knowledge—being the realization of objective truth.[8]

Intellect accounts for various modes of existence and diversities present within the physical and metaphysical world and held by the modicum of peoples, groups, societies, religions, and cultures. It is achieved through a particular pedagogical process whereby those virtues are transferred from teacher to student. It neither accepts beliefs nor claims within or against peoples not rooted in – or at least inconsistent with – a (metaphysically and epistemically) realist portrayal of existence. In this way, the starting point for our presentation of *Intellect* is the promotion of objective truth; moreover, this includes the possibility of religious truth. The two Finnisian values of religion and knowledge can work in conjunction, if the religion in question is open to reason and faith e.g., Catholicism's well-known adage of "faith and reason" (*fides et ratio*) – see John Paul II's *Fides et Ratio* (2000).

This paper proceeds in six subsequent parts. Part I characterizes education and virtue epistemology as outlined in the salient historical and contemporary literature. Part II introduces Sen's conception of capability as the basis for welfare. Part III looks at the capability of education to place it within the framework of *Intellect.* Part IV presents *Intellect*'s underlying virtues. Part V responds to a possible relativist critique of *Intellect*. Part VI concludes. Alongside the promotion of realism about truth, this paper's goal is to implement an *Intellect*-based model within a practical framework to improve the quality of education in both developed and developing country curricula.

Education and Virtue Epistemology

Tangible skills relate to reading and understanding words, calculating numbers, measuring distances, and observing movements in objects and changes in their physical characteristics related to colour, size, shape, texture, and state. The skills are also applicable to understanding extrinsic events, both past and present. In general, skills regard a sensory understanding and hence have their proper object sensory knowledge related to concrete and practical ends. Intangible virtues cannot be directly perceived by the senses – though we might say they are 'perceived' by the intellect.[9] Examples include the implantation of qualities such as compassion, self-confidence, and justice, which are – and ought to be – salient to our understanding of interpersonal relations and societal structures. The inculcation of these virtues does not necessarily relate to learning a positive skill such as reading, writing, or calculation; nor are there set methods or guidelines for these virtues to become rooted within an individual. Of course, each virtue has parameters, but these parameters are specified and defined by the definition of the virtue itself rather than an explicit methodology of choice e.g., being 'just' requires giving to another what they are owed (parameter), but there is no context-less universal rule for implementation (method) by which this is achieved.

Virtue, an active, developing, reliable characteristic/disposition of a human person,[10] while being intrinsically beneficial in raising the individual in the formation of her character, is instrumental because it informs one's worldview or *Weltanschauung* while also shaping the interpersonal relationships and societal structures within which she lives. Here, we use 'character' as defined by Jason Baehr (2017, p. 1153): "…a person's character is comprised of her dispositions to act, think and feel in various ways". We will take Annas' definition of virtue and Baehr's broad Aristotelian outline of 'character' to be correct inasmuch as these provide a holistic and excellence-centred definition. As an intrinsic matter, virtues are good to have in and of themselves despite their practical ends. They are not desirable character-qualities for any subservient ends – even if there are subservient, useful ends. The argument for this claim is found in Finnis: being clear-headed is more desirable to being muddle; qualities of compassion and justice are superior to shrewdness or injustice. In other words, from the mere desirability of the virtues against their counterpart vices, it follows that virtues are desirable qualities in and of themselves. What is implied in contemporary educational paradigms of the sort we describe in this paper is a *tacit commitment* to John Finnis's (1980) natural law conception of knowledge, which is a basic value aimed at achieving truth. Education assumes that we teach truth, whether in the form of tangible or intangible skills. We are not endorsing Finnis' natural law theory in every respect, but we are using his language of basic goods because that is how contemporary education treats the value of education, namely, as a basic, truth-oriented good that at least partially fulfils human nature.

Despite their conceptual differences, Annas points to the relationship between intangible virtues and tangible skills. She states, "…[virtue] shares the intellectual structure of a skill where we find not only the need to learn but the drive to aspire, and hence the need to 'give an account', the need for articulate conveying of reasons why what is done is done" (Annas, 2001, p. 20). It is consistent with our account of both tangible skills and intangible virtues to consider virtue to be a specific kind of skill, that is, a non-tangible skill. In the pre-modern world, specialized institutions of knowledge were available to the societal elite or those who demonstrated particularly high academic aptitude[11]; however, the general shift of education across nation states has been to make tangible skill learning the purview of formal educational institutions[12] – though our argument in no way diminishes the household's role in education.[13]

Sen's Capability Approach to Welfare

In *Development as Freedom*, Sen expressed his dissatisfaction with the Rawlsian conception of equality as being the normative demand of a person's condition and primary goods as being the metric by which to measure utility. For Sen, people are diversely situated such that each person requires a different amount of primary goods to satisfy similar needs. Moreover, Sen differentiates between capabilities and functionings. He states that "[a] person's 'capability' refers to the alternative combinations of functionings that are feasible for her to achieve. … While the combination of a person's functionings reflects her actual *achievements*, the capability set represents the *freedom* to achieve" (Sen, 1999, p. 75). Capabilities are the distinct options available while functionings are those capabilities the individual actually achieves. Capability thus consists of the various combinations of functionings, which Aristotelians conceive as the building blocks for 'flourishment' or *eudemonia (*Aristotle, 1998).[14]

Sen gives three reasons for favouring a capability rather than Rawlsian or Utilitarian approach: 1) it concentrates on those deprivations *intrinsically* important to the individual (unlike low income, which is only *instrumentally* significant; 2) there exist influences on capability deprivation (and subsequently on real poverty) other than low-income levels; and 3) low income is circumstantial within and between communities, families, or even to different individuals. While recognizing that Sen's conception of welfare as a function of capability remains incomplete, we concede that his reasons adequately address the superiority of the capability approach compared to previous approaches.

We agree with Sen's first reason for championing the capability approach because deprivations vary according to each individual. The value of education or health care may differ according to an individual given her particular age, location, or aspirations. Likewise, income serves as an instrument in attaining primary goods and is therefore useful. A practical example is the Dinka peoples of South Sudan who correlate wealth to the amount of cattle a man owns. In the Dinka culture, a man must possess sufficient amounts of cattle with which to pay a bride's family to marry off his son (Deng, 1998, 2009). Lineage is an intrinsically significant capability for the Dinka that may not be affirmed by other cultures. High income levels—an abundance of cattle for the Dinka—is not demonstrable of welfare itself. Rather, it serves as an instrument to reach welfare.

According to Sen's second reason, a capability approach is also superior. The presence of an independent legal system illustrates as such. Human rights laws, property laws, tort laws, or general legal principles such as the rule of law, which are enacted and enforced pursuant to a legal system bereft of any one overwhelming political (or private) influence provide the capability to live a dignified life. A person can make decisions regarding her private property and have redress against another individual who has allegedly done her wrong, irrespective of income level. In such a system, income levels do not play a factor in attempting to exercise a civil or political right. In Sen's framework, this represents a substantive freedom.

Finally, Sen's third reason is fulfilled given that individuals variably prioritize income according to individualized desires or aspirations. An academic motivated by scholarship or discovery or a public official dedicated to the service of her constituents may place less value on income. Conversely, a business executive expanding her business or a consumer wishing to increase her material goods will preferentially value a high income. At the communal level, the Dinka example is still useful. The Dinka peoples place a high value on non-monetary qualities such as respect for others (*atheek*) and social order (*cieng*). Low income is not illustrative of poverty to the same extent as a lack of *atheek* or *cieng* (Deng, 1998). Therefore, capability can provide an expansive understanding of welfare that considers both intrinsic and instrumental components.

The Capability of Education

As discussed above, our focus is on education as the process by which knowledge is transferred from teacher to student. Education aims to actualize the ability to seek and attain knowledge to understand both the physical world and metaphysical realities that inform interpersonal relations, societal structures, and one's comprehensive worldview. This capability is contingent upon the availability of physical institutions, materials, teachers, parents, proper nutrition, proper standards of living, and private and public funds. Likewise, education is contingent upon intangible factors such as societal and familial support and encouragement, an impetus to learn on the student's part, time, energy, and dedication on the part of all parties as well as a general societal focus on education's centrality to the well-being of the individual and community. (See *e.g.*, Metcalfe, 2013)

Education appears as a unique functioning pursuant to Aristotle's (2002) observation that "all human beings by nature desire to know".[15] From an Aristotelian standpoint, the pursuit of knowledge through the capability of education becomes a functioning whenever the capability is present. In contrast, as Sen rightly notes, capabilities other than education may not necessarily be converted into functioning. For instance, despite the capability to procure sanitary nutrition, a person may choose to fast thereby not allowing this capability to turn into the functioning of satiation. Likewise, the capability of political participation within a democratic system may be fully available to a citizen yet she may choose not to exercise such a capability to convert it into its corresponding functioning (Sen, 1999).

In its broadest sense, the capability of education as both skill and virtue transfer is predominantly utilized given its availability. To function as well as understand the world, a student will seize the opportunity to be educated if one is provided. The capability of education can become a functioning without the student-child even knowing it. Virtues can be imparted, or skills can be taught through habitual processes that render a student a passive consumer. As opposed to other capabilities that may not be exercised to become functionings, education, more often than not, will be converted into a functioning given the modes and methods whereby the conversion can take place.

For Finnis, the instrumental form of knowledge means that its acquisition serves a distinctive goal to benefit an individual and/or society through learning a skill or practice. A societal benefit can include a profession or trade or a wherewithal to participate fully within political or economic processes. Finnis construes curiosity-based knowledge – distinct from knowledge *via vana curiositas* – as a virtue simply because "[*i*]*t would be good to find out*" (Sen, 1999, pp. 60-61). (Of course, one might add *ceteris paribus* clauses here cf. ff50). In a word: being well-informed is preferable to being muddled (Sen, 1999). Finnis further elaborates on this type of knowledge by noting that not all things are equally worth knowing and not every form of learning is equally valuable (Sen, 1999). This notion relates to *Intellect* which, as an avenue to optimal welfare, places priority on certain virtues over others in order to inculcate the basic value of knowledge. *Intellect* serves as a curiosity-based knowledge since it instils the virtues of clear-headedness and a state of being well-informed. *Intellect* is also instrumental since being clear-headed and well-informed enables the individual to hone her perspective to understand in greater depth and accept the truth of the matter in question. Finally, education as a precursor capability to acquiring the Finnisian basic value of knowledge and then Sen's conception of welfare is a holistic process. It teaches both tangible skills while inculcating intangible virtues that consider immaterialities.

Education as *Intellect*

In the *Nichomachean Ethics*, Aristotle outlines the most important experiential spheres. He accords each sphere a corresponding normative virtue. For summary of the spheres and virtues, see Nussbaum & Sen, 1993, pp. 246-27). For Aristotle, there is a correct choice of response within every sphere (Nussbaum, 1993).[16] '*Intellect'* is taken from Aristotle's terminology of the 'intellectual life' sphere. The following, in no particular order, is a list of normative virtues that ought to be inculcated through the formal educational process. This list is of our own design, and we define each virtue with illustrations and definitions. These virtues define *Intellect* as a normative *midfare*—the non-utility function achieved through a primary good, in this case education. As we outline them, we will define, explain, and argue for each individual virtue in the context of a student's education.

Creativity

Creativity is the ability to generate innovative ideas, concepts, methods, or technology. These ideas can be independent or dependent upon previous knowledge, concepts, and conceptual systems. Creativity allows the student-child to use her intellectual capacities to envision things differently than the present. By having and utilizing creativity, the student uses her imagination such that she and those around her can view the world anew. This virtue is essential within education. It allows the student to use the full capacity of her mind to create new or modified knowledge irrespective of how it affects current knowledge 'holders,' ideas, concepts, or any paradigm contingent upon current understanding. Creativity enables the student to be iconoclastic when there exists a superior substantive or procedural mechanism, theoretical or practical, to those currently available. Conversely, if creativity is not nurtured, the student is entrenched in a malaise where her potential goes unfulfilled and the world cannot be exposed to a full extent of knowledge that can possibly be conceived and applied.

Curiosity

Curiosity is a desire to know something.[17] This is similar to the Finnisian conception, described above. While the term is the same, the virtue of curiosity is distinct from Finnis' classification of curiosity-based knowledge, which he sees in contradistinction to instrumental knowledge. As a virtue, curiosity is an intrinsic quality generated and nurtured within the student where she has the desire to further comprehend any realm of enquiry. Curiosity (as we use it) is a functionally insatiable appetite to continually improve one's current level of knowledge. Curiosity can relate to understanding the physical world through the hard sciences, the societal world through the social sciences, or the aesthetic or metaphysical world through the arts and humanities. Curiosity is also inextricably linked to the knowledge transferor whose own curiosity is a factor in the quality and quantity of curiosity instilled within the student.[18] If institutions do not cultivate curiosity amongst their students, they pose a core educational deficit against them by limiting possibilities of intellectual inquiry and knowledge.

Tolerance

Tolerance is the disposition of permitting, allowing, or accepting other beliefs and actions. In education, it allows one to accept realities different from which she is accustomed.[19] Tolerance instils acceptance of individuals or societies with different beliefs or practices. From all of the virtues of *Intellect*, tolerance may be the most controversial since it does not accept beliefs merely based on opinion[20] with no epistemic basis. Some opinions should not be tolerated – especially those endorsing intrinsically immoral or evil actions. For example, female genital mutilation is an immoral practice that does not warrant tolerance.[21] Educational institutions have a responsibility to cultivate tolerance in their students inasmuch as they are often public institutions with pluralist values, even if there is a baseline moral uniformity.[22]

Humility

Humility is an honest look in the mirror: a simultaneous affirmation of what one is and negation of what one is not - no more, no less. In relation to *Intellect*, humility is linked to a cautious behaviour that instils a sense of understanding of one's own limitations and hence one's possible limited achievements. As a virtue, humility reminds the student that acquiring knowledge is infinite. In the writings of St. Thomas Aquinas, he repeatedly emphasized that humility was not mere shyness or not appreciating one's gifts and talents. Rather, it was ordered towards the virtue of magnanimity; that is, the striving for the higher things. (See *e.g.*, Aquinas, 1882). Boyd (2014) specifies humility by arguing that it:

(1) Operates according to right reason,
(2) Knows the place to which its possessor has been assigned,
(3) Provides a restraint on the desire for honor, and
(4) Can function appropriately only with its twin virtue of magnanimity.

As we understand the conjunction of (1)-(4), humility is the sober-headed, objective appraisal (whether evaluatively good or bad) of one's character and the recognition of what it ought to be. If humility is not fostered as a virtue in educational institutions, there is a risk of intellectual pride, an inability of self-correction, listening of others and the possibility of meaningfully existing within an intellectual community.

Integrity

Integrity is wholeness or completeness of character, a consistency of inner feelings and outer action. It has been described as what a person does when no one else is watching (May, 1984). As instilled through *Intellect*, this virtue inculcates the notion that the values by which a student lives her life should be (or aim to be) consistent in all circumstances. In education, a proper inculcation of this virtue ensures that a student will be honest in her acquisition and/or practice of knowledge and not act with any malicious intent. This virtue's pragmatic effect is palpable. Individuals part of a society, and educational institutions generally rooted in integrity will be unlikely to lie, cheat, or defraud one another as such occurrences require an inconsistency of inner and outer feelings or, otherwise, a lack of wholeness in character.

Impartiality

Impartiality is an attempted absence of implicit or explicit bias when reviewing competing perspectives. Impartiality is another virtue related to *Intellect*'s preference for the truth. Impartiality, like the classical virtue of prudence or *prudentia*, ensures a sobriety and even-headedness when making decisions. The driving force of impartiality is the desire for the truth of a matter rather than a subjectively desired outcome. The impartial student bases her views on how things are rather than accepting skewed presentations by others. Educationally speaking, a student entrenched in the virtue of impartiality is difficult to persuade by socio-political manipulation, propaganda, and emotion.

Diligence

Diligence is conscientiousness in performing a particular task. In education, the diligent student is one who studies tangible skills such as reading, writing, numeracy, the arts, and sciences with a focus on detail. This virtue relates to curiosity and humility. The diligent student will continue to yearn for knowledge and will understand that the totality of knowledge can never be encompassed by a mere human being. The diligent student values both the time and effort required to adequately learn a skill or concept, whether attained for intrinsic or instrumental purposes. Educational institutions have a responsibility in cultivating the virtue of diligence in their students. When it is not fostered, the result can be unprofessional, hasty, and premature work. The concern here parallels (or is found in) a more recent debate on whether graduate students should submit their papers to professional journals since their work may be (and usually is) at its infancy. At any level of education, diligence is the preventative prognosis against intellectual prematurity. The diligent person is able to recognize her contributions to a specific intellectual investigation, identifying her own methods, presuppositions, limitations, and relation to the broader intellectual community.

Determination

Determination is a resolve to fulfill a purpose from the outset of a task despite hardships, obstacles and, in idiomatic language, the grunt work of one's task. In education, while it can relate to acquiring knowledge for its own sake, it pertains more to the instrumental notion of education that has a purpose other than the pursuit of knowledge. This virtue instils a sense of drive and ability to persevere upon receiving an initial negative result. Determination relates to diligence and humility given that knowledge can be pursued to infinite depths. The determined student is also curious because she seeks to know something of which she is currently ignorant. If educational institutions do not invest in instilling determination, the results can include incomplete work, failure to meet expected educational standards, and an overall lack of intellectual self-confidence. As a virtue-based approach, *Intellect* does not necessarily limit or even define the substance or methodology of educational institutions. In fact, *Intellect* encourages novel substantive and methodological knowledge if it improves upon previous knowledge.

Intellect and Relativity

A possible relativist[23] objection to *Intellect* may be to say that its constitutive virtues are subject to variations within and between cultures and peoples.[24] A relativist may exclaim that curiosity or tolerance is acceptable where it does not *offend* higher moralities rooted in an ancient past or religious belief. Others may note that in some cultures, humility is to be minimized given its propensity to instil shyness and thereby inhibit an individual from meeting her potential. Others may yet note that impartiality is subservient to any connection an individual may have to the State or an ethnic or political group. In this light, the relativist views the purpose of education as contingent upon a student's contextual and circumstantial reality. To the relativist, the pursuit of knowledge, in the Finnisian sense, is subjective.

An initial issue with the relativist argument is that achieving truth through knowledge becomes but among a myriad of other purposes related to national, ethnic, socio-political, or individual hubris. The prime purpose of knowledge, as envisioned through *Intellect*, is acquiring truth such that other subjective purposes are set aside. While *de facto* variations exist across cultures, *Intellect* views truth as an objective purpose. Knowledge is the means by which to obtain this purpose. Aristotle's characterization of knowledge is thus helpful:

> [y]ou cannot understand anything through a demonstration unless you know the primitive immediate principles. … from perception comes memory, as we call it, and from memory (when it occurs in connection with the same item) experience; for memories which are many in number form a single experience. And from experience, or from all the universal which has come to rest in the soul … there comes a principle of skill or understanding - of skill if it deals with how things come about, of understanding if it deals with how things are (Aristotle, 1994).

As in Aristotle's definition above, *Intellect* emphasizes experience, observation, and reason. Knowledge, in order to achieve truth, must be at education's forefront. Nothing should distract the student from this goal. The relativist's argument merely considers the presence of knowledge if it accords with her subjective beliefs. Generally speaking, the relativist is not concerned with enhancing knowledge according to new methods of discovery and understanding. The relativist remains content with knowledge that accords with her core beliefs while being unable to challenge and eventually alter those beliefs when confronted with new understandings. *Intellect* is not worried about the iconoclastic nature of knowledge. Rather, it welcomes knowledge within its foundational virtues. (To consider further replies to relativism, see Shafer-Landau, 2001, pp. 9-15)

Conclusion

This discussion has presented *Intellect*—a virtue-based approach that serves as the normative *midfare* to education—through Amartya Sen's capability-based conception of welfare. *Intellect* is the set of underlying virtues that prioritize the Finnisian conception of knowledge as the basic value for attaining truth (Finnis, 1980).[26] *Intellect*'s prioritization of truth-seeking through evidence and experiential knowledge in the background, introducing a novel approach that could begin at the primary level and continue up until the secondary level for transitional nations is a worthy endeavour that can assist the next generation of students to reach their full intellectual potential. Given that educational purposes have been driven by secondary ends over truth, character and the development of knowledge, *Intellect* is a rational alternative. If virtues that constitute *Intellect* are applied, there will be a further effort to ensure truth within each circumstance. The quality of and approach to education in a child's formative and adolescent years reflect her basic understanding of the world throughout her life. *Intellect* serves as both a theoretical framework and a practical method by which to approach a student's education. Of course, the underlying assumption behind this framework is that the student will accept and emulate *Intellect*'s virtues. Critiques and improvements to the model presented here are welcomed to produce a robust functioning model of *Intellect* that can be implemented in educational curricula.

Notes

[1] What constitutes the destination of a 'professional' teacher will vary culturally-geographically based on differing evaluative standards. For example, in Canadian educational institutions teachers are officially credited teachers at various levels of education; however, there is no explicit (and relevant) designation of 'professional.'

[2] See e.g., Chesterton (1930): "There begins to be a mere vanity in being educated, whether it be self-educated or merely State-educated. Education ought to be a searchlight given to a man to explore everything, but very specially the things most distant from himself. Education tends to be a spotlight, which is centred entirely on himself. Some improvement may be made by turning equally vivid and perhaps equally vulgar spotlights on a large number of other people as well. But the only final cure is to turn off the limelight and let him realize the stars."

[3] This paper's purpose is not to delineate the (de)merits of the State's role in education. Whether little or great, we are only saying that the State, descriptively, has had *some* role to play in educating its citizens.

[4] Sen's book (1999) is an elaboration upon the lecture by Sen (1979).

[5] The term "functionality" or "functionalities" is traditionally a term attributed to Aristotle even though contemporary authors attribute it to Sen in his definition of capability. Sen acknowledges the Aristotelian roots of "functionalities" in Sen (1999, p. 75).

[6] Throughout this paper, we write *Intellect* italicized and with a capital "I" to distinguish it from any other general characterization of intellect or *intellectus*. When using *Intellect* throughout this article, it will denote

our particular definition of the term that is or might be predicated on a set of normative virtues instilled through the formal education process.

[7] We will not focus on educational institutions' responsibilities in promulgating tradition; however, for further discussion see Pieper (2015, pp. 29-42).

[8] We are using "objectivity" to mean *mind-independent* i.e., the truth-maker of a proposition is mind-independent in the sense that what makes a proposition true or false is *reality*.

[9] For a brief history of virtue epistemology's resurgence in contemporary philosophy, see Zagzebski & Fairweather (2001, pp. 3-14).

[10] For spatial considerations, we cannot defend at any acceptable length our specific definition of virtue; however, this definition and its defence is found in Annas (2001, p. 8).

[11] For a history of the right to education, see Beiter (2006, pp. 17-46).

[12] For tables listing the enrolment in formal schooling over the years, see Countries and Economics (2021).

[13] Recent treatment of adjacent issues is found in Abbarno (2020).

[14] In the *Nicomachean Ethics*, Aristotle describes *eudaimonia* as the highest good for human beings.

[15] The entire quote is "All human beings by nature desire knowledge. An indication of this is the delight we take in our senses; for even apart from their usefulness they are loved for themselves; and above all others the sense of sight. For not only with a view to action, but even when we are not going to do anything, we prefer sight to almost everything else. The reason is that this, most of all the senses, makes us know and brings to light many differences between things" (Aristotle, 2002).

[16] Some of the spheres that Aristotle indicates include fear (especially fear of death), bodily appetites and their pleasures, distribution of limited resources, management of one's personal property, where others are concerned, attitudes and actions with respect to one's own worth, attitude to slights and damages, and association and living together and the fellowship of words and actions.

[17] For one component of the historical treatment of curiosity (*curiositas*) as a vicious character trait, a related though separate issue, see Rehman (2021).

[18] Concerning the ethical character of teachers and its effect on the student's education, see Lockowski (1997).

[19] To clarify, we are not saying that all morals are indicative of the definitive virtues of *Intellect* – morality is *broader* than virtues, especially on a broader moral nonconsequentialism present in this work.

[20] 'Mythic' in the narrow sense of falsehood, not in the more philosophically accurate sense of a mode of symbolic orientation in the world. See Metcalfe (2013, pp. 1-70).

[21] For discussions concerning female genital mutilation and ethics, see Kopelman (1994), Atoki (1995), Elsayed (2011). Our argument is that as moral objectivists, we are able to distinguish the moral and cultural/social/legal, such that we are able to regard various claims of tolerance as defeasible on *objective* moral grounds.

[22] However, tolerance is a minimum threshold virtue, if taken to mean that we have a moral obligation to habitually cultivate tolerance of others. A helpful distinction may be drawn from Josef Pieper's philosophy of love, where human beings are to be fundamentally *loved*, but some of their (difficult, wrong, etc.) actions *tolerated*. See Pieper (2012).

[23] We do not 'nuance' an approach which balances objectivity and relativism for two reasons. First, our definition of objectivity as mind-independence requires the falsity of relativism. *Even if* our statements about reality are positioned or socially conditioned, we are affirming the thesis that we can still nonetheless know reality as it is. This is the antithesis of relativism. Second, our approach is epistemically open in the sense that statements of 'objectivity' are non-dominating: we take it that knowing reality is a (partially) shared, epistemic endeavour in which dialogue, argumentation and listening are fundamental.

[24] While often the formulation of the relativist's claim (and a shared view of popular anthropologists), this claim is empirically overstated and historically exaggerated. For a comprehensive survey of universally shared natural law moral beliefs *e.g.*, beneficence, justice, duties to elderly and children, mercy, magnanimity, faith, veracity, et cetera, from the writings of ancient Greeks, Romans, Egyptians, Babylonians, Chinese, Norse, Indian, as well as from Christians and Jews, see Lewis, (2001, pp. 731-738). This puts pressure on the relativist to supply the burden of proof against the general moral objectivism of human beings socio-culturally, transnationally, and historically (without recourse to overgeneralizations).

[25] Exceptions do exist to the model in which skill building is the purview of the formal educational institution.

References

Abbarno, G. J. M. (2020). *The ethics of homelessness*: *Philosophical perspectives*. Brill.

Annas, J. (2001). *Intelligent virtue*. Oxford University Press.

Aquinas, Th. (1882). *Opera omnia.* Ex Typographia Polyglotta.

Aristotle. (1994). *Posterior analytics* (J. Barnes, Trans.). Clarendon Press.

Aristotle. (2002). *Metaphysics* (Joe Sachs, Trans.). Green Lion Press.

Aristotle. (1998). *The Nichomachean Ethics* (D. Ross, Trans.). Oxford University Press.

Atoki, M. (1995). Should female circumcision continue to be banned? *Feminist Legal Studies*, *3*, 223-235.

Baehr, J. (2017). Varieties of character and some implications for character education. *Journal of Youth and Adolescence*, *46*, 1153-1161.

Beiter, K. D. (2006). *The protection of the right to education by international law*. Martin Nijhoff Publishers.

Boyd, C. (2014). Pride and humility: Tempering the desire for excellence. In C. Boyd & K. Timpe, *Virtues and their vices* (pp. 245-268). Oxford University Press.

Countries and Economics. (2021). *The World Bank*. http://data.worldbank.org/country

Chesterton, G.K. (1930). The worship of education. *Illustrated London News.*

Cohen, G. A. (1993). Equality of what? On welfare, goods, and capabilities. In M. C. Nussbaum & A. Sen, (Eds.), *The quality of life.* (pp. 9-29). Oxford University Press.

Davidson, T. (1990). *History of education.* Ams Pr Inc.

DeMarco, D. (2000). *The many faces of virtue.* Emmaus Road.

Deng, F. M. (2009). A cultural approach to human rights among the dinka. In W. Twining, (Ed.,), *Human rights, southern voices.* (pp. 44-52). Cambridge University Press.

Deng, F. M. (1998). The cow and the thing called 'what': Dinka cultural perspectives on wealth and poverty. *Journal of International Affairs, 52*, 101-29.

Elsayed, D. M. (2011). Female genital mutilation and ethical issues. *Sudanese Journal of Public Health, 6*, 63-67.

Finnis, J. (1980). *Natural law and natural rights.* Oxford University Press.

Graves, F. P. (1999). *A history of education.* Anmol Publications Pvt. Ltd.

Graves, F.P. (1911). *A history of education before the Middle Ages.* MacMillan.

Halvorsen, K. (1990). Notes on the realization of the human right to education. *Human Rights Quarterly, 12*(341), 341-364.

John Paul II. (2000). *Fides et ratio*. Pauline Books & Media.

Kopelman, L. M. (1994). Female circumcision/genital mutilation and ethical relativism. *Second Opinion, 20*(2), 55-71.

Lewis, C.S. (2001). The abolition of man. In *The signature classics of C.S. Lewis*. (pp. 731-738). Harper One.

May, W. F. (1984). The virtues in a professional setting. *Soundings*, *67*, 245-266.

Mill, J.S. (1991). *On liberty.* Oxford University Press.

Nussbaum, M. (1993). Non-relative virtues: An Aristotelian approach. In Nussbaum, M. C., & Sen, A. (Eds.), *The quality of life*. (pp. 242-268). Oxford University Press.

Nussbaum, M. C., & Sen, A. (Eds.). (1993). *The quality of life*. Oxford University Press

Oderberg, D. (2013). Natural law and rights theory. In G. Gays., & F. D'Agostino, (Eds.,), *The Routledge companion to social and political philosophy*. (pp. 375-86). Routledge.

Parker, S.C. (1970). *A history of modern elementary education.* Littlefield Adams.

Pieper, J. (2012). *On love* (R. Winston., & C. Winston, Trans.). Ignatius.

Pieper, J. (2015). What is, in truth, worth preserving? (D. Farrelly, Trans.). In *Tradition as challenge: Essays and speeches*. (pp. 29-42). St. Augustine's Press.

Rehman, R. (2021). *Sana oculos meos*: Alypius' curiositas in Augustine's Confessiones (6, 8, 13). *Augustinianum*, 61.1, 137-152.

Sen, A. (1999). *Development as freedom.* Oxford University Press.

Sen, A. (1979). Equality of What? The Tanner Lecture on Human Values. Stanford: Stanford University.

Shafer-Landau, R. (2001). Ethical subjectivism. In R. Shafer-Landau., & J. Feinberg, (Eds.,), *Reason and responsibility.* (pp. 9-15). Wadsworth.

Ulich, R. (1945). *History of educational thought.* American Book Company.

Watkins, K. (2012). The power of circumstance: A new approach to measuring educational inequality. Working Paper 5, Centre for Universal Education.

Zagzebski, L., & Fairweather, A. (2001). *Virtue epistemology: Essays on epistemic virtue and responsibility.* Oxford University Press.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 28-40

Artículo/Article

Moral responsability in Plato's philosophy. Commentary on "Timaeus" 86D-87B

Responsabilidad moral en la filosofía de Platón. Comentario sobre "Timeo" 86D-87B

Marc Zapata Pedrosa

Universidad de Barcelona, España
marczapata.trebali@gmail.com
https://orcid.org/0009-0004-2500-729X

DOI https://doi.org/10.48204/2805-1815.8471

INFORMACIÓN DEL ARTÍCULO	ABSTRACT/RESUMEN

Recibido el: 7/4/2025

Aceptado el: 31/8/2025

Keywords:

Human agency, determinism, education, illness, responsibility

Palabras clave:

Agencia humana, determinismo, educación, enfermedad, responsabilidad

Abstract:

This paper explores various interpretations of moral responsibility in Plato's *Timaeus*, addressing the puzzle posed by Taylor on the relationship between determinism and human agency. Four main solutions are analyzed: the denial of authorship, the pedagogical approach, the afterlife theory, and the nuanced causality interpretation; it is argued that all these ultimately fall into the same determinism they aim to resolve. Finally, a fifth interpretative approach, the narrative approach, is proposed, suggesting that the determinism in the *Timaeus* relates to the narrative aspects of the dialogue: *Timaeus*, as a politician, discusses human nature as something determinable by politics, not in an absolute sense.

Resumen:

Este artículo explora diversas interpretaciones de la responsabilidad moral en el *Timeo* de Platón, abordando el rompecabezas planteado por Taylor respecto a la relación entre el determinismo y la agencia humana. Se analizan cuatro soluciones principales: la negación de la autoría, el enfoque pedagógico, la teoría del más allá y la interpretación de causalidad matizada; se argumenta que todas ellas terminan cayendo en el mismo determinismo que intentan resolver. Finalmente, se propone un quinto enfoque interpretativo, el enfoque narrativo, sugiriendo que el determinismo en el *Timeo* se relaciona con los aspectos narrativos del diálogo: Timeo, en tanto que es un político, discute la naturaleza humana como algo determinable por la política, no en un sentido absoluto.

Introduction: Taylor's puzzle

The reconciliation of the diverse doctrines found in Plato's dialogues has been a central concern of Platonic studies since their earliest days, dating back to the lifetime of the Athenian master himself, as it is pointed out by Aristotle. Among the controversies surrounding the differing views expressed in his works, Taylor (1927) highlights a debate between the *Timaeus* and the rest of the Corpus on the issue of moral responsibility. In his monumental Commentary, He argues that, in the final sections of the *Timaeus*, the eponymous interlocutor presents a theory that challenges and undermines the achievements of Socratic-Platonic philosophy in this domain:

> If we read T.'s exposition of it carefully, I think we should be struck by a curious fact. His exposition explains away that very fact of moral responsibility on which Socrates, Plato, Aristotle and Timaues himself, when he is talking ethics and not medicine, are all anxious to insist. The interpretation he proceeds to give of the formula is therefore non-Platonic and non-Socratic. (Taylor, 1927, p. 611)

This is because, in Taylor's view, the genuinely Platonic doctrine of moral accountability is defined by a strong defense of each individual's responsibility in the process of self-formation. Awakening this sense of responsibility, or care for oneself, would appear to be, in his own words, the only intention of the "gadfly of Athens," as recorded in the *Apology* (Plato, 1997):

> For I go around doing nothing but persuading both young and old among you not to care for your body or your wealth in preference to or as strongly as for the best possible state of your soul, as I say to you: Wealth does not bring about excellence, but excellence makes wealth and everything else good for men, both individually and collectively. (30a-b)

One must not fall into the notion that this is a purely Socratic idea, later abandoned by Plato during the intellectual emancipation evident in his dialogues of maturity and old age. In fact, it could be argued that this very notion is rediscovered in the Myth of Er, which, almost as a culmination of the argument presented in the *Republic*, separates human life from divine interests, rendering the human being responsible for their own destiny:

> A demon will not choose you, but you will choose a demon. Let the one who is drawn first by lot select a way of life, to which he will necessarily be bound. As for

> excellence, it belongs to no one, but each will have a greater or lesser share of it depending on whether they honor or despise it. The responsibility lies with the one who chooses; God is free from blame. (Plato, 1968, 617e)

However, it is true that the *Timaeus* presents a radically opposed perspective on this very matter. In the final sections of the cosmological dialogue, when the discussion turns to the maladies of the soul—which are nothing other than what Aristotle would later call vices—the principal interlocutor asserts that all evil afflicting the human soul does not stem from any particular activity or conduct of the individual. Instead, it arises from an asymmetry between soul and body, i.e., biological determinism, or a deficiency in the educational process, i.e., social and educational determinism:

> No one is wilfully bad. A man becomes bad, rather, as a result of one or another corrupt condition of his body and an uneducated upbringing [...] that is how all of us who are bad come to be that way—the products of two causes both entirely beyond our control. It is the begetters far more than the begotten, and the nurturers far more than the nurtured, that bear the blame for all this. (Plato, 2000, 86d-87b)

In this way, Taylor constructs a puzzle: the challenge of reconciling the Socratic-Platonic notion of responsibility—characterized by attributing to humans the responsibility for their destiny and their active participation in shaping their moral character—with the Timaean notion, which absolves humans of such responsibility and situates morality within the domains of biology and education. From my perspective, the central issue of this puzzle lies in the differing roles that both theories assign to the individual in self-formation. The Socratic view of responsibility, despite adhering to the typically intellectualist premise that no one intentionally does evil, nonetheless seems to encourage the individual to take charge of their own education. The Timaean view, by contrast, appears to succumb—under the same premise—to the idea that an individual's life is entirely conditioned by the nature of their body or the city responsible for providing their education. Thus, in the first of these theories, the weight of morality ultimately rests on the individual, whereas in the second, it shifts to external causes, such as nature and politics. To address this issue, the proposals of various scholars will be evaluated, and, ultimately, a suggested solution will be presented.

The attempts of solution (and its problems)

In response to the puzzle, he himself poses, Taylor naturally proposes a solution, which we shall call the "denial of authorship approach." In his view, the contradiction between the *Timaeus* and the rest of the *Corpus* is not doctrinally problematic, as it belongs to *Timaeus*, not to Plato. Thus, the contradiction between the various dialogues would be caused by contemporary academia, which erroneously attributes the words of Plato's characters to the philosopher himself. However, this attempt to preserve doctrinal harmony among the Platonic texts is, in my view, a trap—a supposed harmony that, in reality, rests on the denial of a Platonic doctrine. This is because, if one accepts this interpretative principle and applies it consistently to the Parmenides, Sophist, Statesman, and other dialogues where Socrates is not the main interlocutor—even to those where he is—we would be forced to conclude that we lack a single definitive opinion from Plato. This would dissolve the Corpus into a mere doxographical collection and render the study of Platonic philosophy meaningless. Consequently, in attempting to safeguard the unity of Platonic doctrine, this approach ultimately dissolves it into a collection of disparate opinions.

That said, pointing out that the consequences of a particular interpretative approach are undesirable does not, in itself, constitute a solid argument against it. It could well be the case that Plato merely collected the opinions of the philosophical elite of his time, that the dialogues were nothing more than a portrayal of the intellectual landscape of his era, and that, ultimately, there is no such thing as a Platonic doctrine. Certainly, his texts provide no explicit evidence to the contrary. Nevertheless, in my view, there are various pieces of evidence that demonstrate that the *Timaeus* contains what could be described as genuinely Platonic philosophy. This would make it illegitimate to simply attribute the ideas presented in the dialogue to the statesman *Timaeus* alone.

The first of these arguments is found in the text of the *Timaeus* itself, more specifically in [....]. During the exposition of the myth concerning the genesis of the universe, *Timaeus* states that the demiurge creates the world by observing the Forms and using them as a model. In my view, this doctrine is clearly related to what is presented in the *Republic* and, as Hill (2016) points out, also to the Phaedrus. Therefore, even if we cannot take *Timaeus*' words as those of Plato himself, we can assume they are imbued with a strong dose of Platonism. It could naturally be argued that it was perhaps Socrates himself who, in the conversation from the previous day recounted at the beginning of the dialogue (17a–19a), explained the Theory of Forms to *Timaeus*. Since this section recalls the content of the *Republic*, it suggests a connection between this dialogue and the *Timaeus*, and it could be supposed, therefore, that *Timaeus* might have learned about Platonic metaphysics through this means. However, the truth is that when the previous conversation is recalled—which is undoubtedly linked to the *Republic*—it only refers to its political content, without mentioning the Forms or any other concept specific to Platonic metaphysics. Thus, while we know that *Timaeus* is well-versed in the Forms, we do not know when he became familiar with this doctrine. For this reason, it is plausible, in my opinion, to assume that a philosopher who so masterfully employs the concept of eidos is not a statesman unknown to historical tradition, but Plato himself.

The second argument against this interpretation is found in the immediate reception of the *Timaeus*, that is, in the philosophy of Aristotle. The Stagirite was directly familiar with the text and even cites it on multiple occasions throughout his work, yet at no point does he question the attribution of its ideas to his teacher, Plato. Significantly, in *Physics*, IV, 2, when reflecting on the concept of *tópos* (place), he refers to the *Timaeus*, stating the following: Even so, one would have to ask Plato —if it is necessary to make a digression— why forms and numbers are not in a place, considering that place is participatory, whether it be that the participatory is the great and the small, or matter, as he has written in the *Timaeus*. (Plato, 2000, 209b 33-35)

Beyond the strictly philosophical content of the passage, what is relevant for our work lies in the fact that, in Aristotle's view, Plato takes ownership of the doctrines presented in this dialogue. Similarly, in On Generation and Corruption, Aristotle states that Plato 'establishes,' 'said,' and 'analyzes' various theories from the *Timaeus*. Considering the limited testimonies that have reached us regarding the functioning of the Academy and the relationship between disciple and teacher, it is difficult to think that, if the doctrines of the *Timaeus* did not actually belong to Plato, Aristotle would have expressed himself in this way, attributing to the master doctrines that were not his own. Therefore, I believe it is reasonable to accept Aristotle's testimony in this case and conclude that the doctrines presented in the *Timaeus* do not belong to an unknown Greek statesman but are indeed a display of genuinely Platonic philosophy. As a result, Taylor's attempt at reconciliation is ruled out, and we must therefore continue questioning the reconciliation between the various Platonic dialogues.

In this regard, it is interesting to bring up Sedley's (2019) article, in which he argues in favour of the thesis that the *Timaeus* is truly a vehicle for the most strictly Platonic ideas, despite the apparent contradictions that may arise. In this way, we could affirm that we are aligned with Sedley, insofar as we also consider that the *Timaeus* contains the essence of Platonic philosophy.

Some scholars have attempted to overcome this contradiction through what we will call the pedagogical approach, given that this interpretation emphasizes the role of education in the moral development of man, to the point of, in my view, falling into a form of pedagogical determinism, where a man's morality depends entirely on his education; broadly speaking, this interpretative approach is characterized by the assumption that moral evil originates from the pre-demiurgic material substratum and that it is possible to overcome it through education. Among the scholars who could be placed within this perspective is Gill (2000), who acknowledges the puzzle hidden in the text at hand and seeks to solve it by relating the passage to other key points in the Corpus. In his view, the description of the soul's diseases as a consequence of certain physical defects complements the Socratic-Platonic maxim that no one does evil voluntarily, but rather that evil is always linked to a certain kind of ignorance. In the case of the *Timaeus*, what Plato would be attempting to show is that ignorance regarding the functioning of the body is the foundation of the soul's evils, insofar as it prevents proper care of the body. Therefore, it is not that human beings are determined by their bodies, but rather that, by being unaware of how the body should be treated, they lose control and become the source of various diseases.

This interpretation, although it seems to resolve the problem of determinism, can be criticized when considering the literal meaning of the text in question. Gill, for his part, claims that ignorance is the cause of neglecting the body, which leads to the soul's diseases; however, in the *Timaeus*, it is stated that ignorance is a consequence—that is, something that follows this loss of control.

Therefore, it is true that Gill's interpretation harmonizes the various dialogues, but in my view, it does so at the expense of the literal meaning of the *Timaeus* text—or, in other words, it projects the doctrine of other dialogues onto the text at hand. In the *Timaeus*, ignorance is the result of a bodily disorder, not its origin.

Within this same pedagogical approach, we find Pears (2015), who argues that the contradiction between the cosmological determinism of the *Timaeus* and the phenomenon of human freedom can be overcome by referring to the notion of "progress." In this way, although human beings are in some way conditioned by the material substratum from which they are composed, they can progressively rise above this state through education—provided that this education consists of aligning the various parts of the soul with the harmony of the cosmic revolutions established by the demiurge. However, Pears himself, in the conclusion of his article, acknowledges the need to further explore this interpretative line, as it does not fully resolve the issue.

This position is close to that of Campbell (2020), who believes that a careful reading of the text reveals that the origin of human moral evil lies solely—this nuance of singularity in the origin is the main point Campbell emphasizes—in the human bodily condition, which, nevertheless, can be corrected through pedagogy. Steel (2001) also attributes the origin of evil to a physical issue related to the pre-demiurgic state of matter. However, instead of locating it in the body, he identifies the passions as the cause of disease, and moral development would involve educating them.

In this way, these five authors indicate that moral evil originates in matter, the body, ignorance, or the passions—all of which are derived from the pre-demiurgic state of matter, which the god cannot shape entirely at will. As a consequence, human beings are imperfect. For the matter at hand, all these concepts are analogous, as the four proposals locate the origin of the soul's ailments in physical issues and their remedy in pedagogical methods, which, moreover, are presented within the text itself. Therefore, not only would there be no biological determinism in the *Timaeus*, but the text itself would also provide the keys to overcoming it by emphasizing the importance of education in human development.

However, this approach still does not resolve the problem raised by Taylor; in my view, it merely avoids it. Education, as these authors conceive it, is necessarily an external stimulus, separate from the moral subject itself, insofar as humans are conditioned by their pre-demiurgic matter. In this way, the fact that a person can emerge from their state of soul sickness through the education provided by teachers or elders does not imply that they become responsible for their actions. Rather, it extends the scope of determinism to the realm of education. Thus, the pedagogical approach assumes that we are biologically determined, though not absolutely so. Through education—which does not originate from the individual (since they are biologically prone to disease) but is instead provided by the city—it becomes possible to prevent these ailments and improve individuals' moral character.

To fully resolve the puzzle posed by Taylor and align the doctrine of moral responsibility in the *Timaeus* with the rest of Plato's dialogues, it would be necessary for the pedagogical approach to demonstrate how a human being—in their material and bodily condition, which is, in this view, imperfect and prone to ignorance—is capable of actively participating in their own education. Otherwise, we must admit that education is only possible within the framework of political life and that the moral character of individuals is merely the product of their educators' influence. Consequently, it would not be possible to speak of moral responsibility in the strong sense. Instead, the pedagogical approach would lead us into a form of pedagogical determinism, where responsibility rests not with the student but with the teacher, as is indeed suggested in the text (Plato, 2000, 87b). Therefore, insofar as the pedagogical approach fails to harmonize the notion of responsibility in the *Timaeus* with that of other dialogues—and, on the contrary, deepens the pedagogical determinism that seems to emerge from the text, which contradicts the rest of the Corpus according to Taylor—it must be ruled out as a solution to his puzzle.

To resolve the puzzle, one could turn to the interesting article by Kamtekar (2016), in which she explores what could be considered a third way for our problem, which we will call the afterlife approach. This approach emphasizes moral responsibility through the concepts of reincarnation and punishment presented in the *Timaeus*, thereby linking this dialogue with the *Laws*, *Gorgias*, *Phaedrus*, and, of course, the *Republic*. In this way, the author highlights the fact that reincarnation, which is described in the *Timaeus* as a punishment, would form part of a process of moral progress orchestrated by the gods to establish the most perfect possible harmony in the cosmos. In the same vein, we find the article by Stalley (1996), which emphasizes the role of punishment in Platonic cosmology—specifically, the role of punishment and moral development in the demiurge's plan for cosmic harmony.

However, it is important to note a significant issue that arises when considering this approach. At the moment of reincarnation, the human being is judged for their conduct on Earth, meaning that the afterlife is conditioned by their earthly life. Yet, it should not be forgotten that this is also conditioned by the biological determinism described by the *Timaeus* itself. Ultimately, situating the moment of moral responsibility in the judgment after death, without first addressing the problem of biological determinism, far from achieving the supposed moral progress that Kamtekar and Stalley point to, could lead to the opposite: from a bad bodily condition comes bad conduct, and from bad conduct comes a bad reincarnation, which consists of a bad bodily condition, and so on, resulting in a clear moral degeneration in which the individual cannot take responsibility. Thus, this approach does not solve the problem of determinism; rather, it exacerbates it, now involving divine action in the moral development of the human being. As we have pointed out, this had been overcome by Plato in the *Republic*.

Furthermore, placing human moral progress within the demiurge's plan, rather than overcoming the determinism we are facing in this article, ultimately exacerbates it, since this moral progress would be the result of divine will rather than human agency. Consequently, Kamtekar suggests that, in her interpretation, the gods and humanity share responsibility for human beings. For these reasons, we believe that the afterlife approach is not suitable for overcoming the determinism pointed out by Taylor.

The last approach we will consider in this article is that of Jorgenson (2021), who attempts to overcome the problem of determinism in the *Timaeus* by suggesting that the theory of causality in the text should not be understood in a strong sense but rather requires nuance. For this reason, we will refer to this interpretation as the nuanced approach. Jorgenson addresses the problem of responsibility in dialogue with Taylor and Gill and points out that both authors err in assuming that the notion of causality in the *Timaeus* is linked to moral responsibility. Due to this mistake, they assume that, when *Timaeus* says that parents are the cause of their children's evil, it implies that they are therefore responsible for it. Jorgenson, on the other hand, argues that this idea of causality is morally neutral, and therefore, while parents are the cause of the evil in their children, this should not be understood in a deterministic framework. Rather, it is simply pointing out the origin of the evil, assigning the children the task of overcoming the limitations imposed by their parents in conceiving them. it is not merely that we cannot be blamed for our badness, but that the notion of blame itself makes little sense, at least at this lofty level of analysis. This point seems to be missed by most commentators, who take *Timaeus* to say that blame is transferred from the child to the parents. In fact, there is a subtle, but significant difference in meaning between the words *Timaeus* uses in the two cases. Initially, he says that no one does wrong willingly and hence that those who are bad are “wrongly blamed” (*ouk orthōs oneidizeitai*, 86d7) for their actions. But when he attributes responsibility to parents and educators, he uses not *oneidizeitai*, but *aitiateon*. The latter term can mean “to blame” someone for a fault, which is how it is generally interpreted here, but it also has the more neutral sense of “identify as the cause.” (p. 270)

This interpretation, while quite interesting, reveals at least two weaknesses. The first is that the philological exercise proposed by Jorgenson rests on a single use of *aitiaeton* detached from its moral sense. In fact, the author bases his interpretation on a passage from the *Republic* (397c2-7), where Socrates distinguishes between being responsible (*aitios*) and being identified as the cause (*aitiaeton*). However, the fact that this is the only example the author presents in his study compels the reader to remain cautious, awaiting further philological studies to shed more light on the matter. Certainly, Jorgenson's observation is of great interest; however, the evidence in its favor is scant.

The second issue is that alternative lines of interpretation regarding the meaning of *aitiaeton* can be traced. In fact, the term is a verbal adjective derived from *aitiaomai*, meaning "to accuse," a usage found in the *Republic* (562d), which fits perfectly with the sense of the *Timaeus* passage. Therefore, while Jorgenson's interpretation is genuinely intriguing, it warrants a deeper study of Plato's terminological usages, which, for now, has not been fully resolved.

In summary, in this section we have evaluated four possible solutions to Taylor's puzzle about moral responsibility in Plato's philosophy. The first, the route of denying authorship, proved insufficient when considering various ancient testimonies about the

Timaeus, among other arguments presented. The second, the pedagogical approach, was rejected on the grounds that it failed to escape determinism, instead shifting it to the realm of education, which does not resolve our puzzle. The third, the afterlife approach, was found to be contrary to moral progress and additionally involves the gods in the fate of humans, leading to theological determinism. Finally, the fourth and last, the nuanced approach, while interesting, was considered problematic due to the lack of philological evidence in its favour, on one hand, and the evidence against it, on the other.

Conclusion: Cornford's solution (and a contribution to it)

Cornford (1937/1997), who outright rejects Taylor's deterministic interpretation, argues that the passage in question should be understood in light of the narrative elements that constitute the *Timaeus*. For this reason, we will refer to this interpretative approach as the narrative approach. In Cornford's view, it is crucial to consider, first, that this is a dialogue on natural philosophy, and its exploration of ethical issues should be interpreted from this perspective. Second, it must not be overlooked that *Timaeus*, the character, is a statesman and an expert in both politics and astronomy. Thus, Cornford's interpretation emphasizes that what *Timaeus* says should not be understood as an absolute assertion of the determined nature of human beings. Instead, it merely establishes the framework of natural determination that a politician must consider when performing their duties. In other words, *Timaeus* is simply articulating what a statesman needs to understand about natural philosophy for the proper execution of their political responsibilities.

In his view, considering the doctrine of the *Timaeus* in this way not only resolves the alleged dissonance between this dialogue and the rest but also makes it legitimate to establish a concordance with what is presented in the Laws. Indeed, in this dialogue, the description of disease only makes sense in relation to the subsequent explanation of its methods of prevention:

> The doctrine of the Laws is in harmony with our passage. The evils here described are to be pitied because their origin lies in causes at work when a man cannot have begun to exercise rational control, and they are remediable if taken in hand before he comes 'totally and obstinately wicked'. This is the answer to the criticism that *Timaeus* leaves out of account 'real wickedness' and 'conceive of no wickedness that is more than weakness'. The passage is not concerned with the ingrained and

> irremediable vice which calls for punishment or extermination. A physical treatise may confine itself to hygiene. All that is needed is the mild preventive remedies described in the next paragraphs. (p. 349)

In this context, if we consider that *Timaeus* is not describing human nature as something fixed and predetermined, but rather as something shaped and influenced by the political system in place, it opens up an interesting perspective on the role of society in human development. This view suggests that individuals are not simply bound by innate characteristics but are instead moulded by the values, structures, and norms of the political environment they inhabit. Such an interpretation emphasizes the adaptability and malleability of human beings, making political and social frameworks key factors in defining human behaviour and ethical principles. Therefore, *Timaeus* seems to present a dynamic vision of humanity, one that evolves and responds to the conditions imposed by governance and social organization.

At this point, it is worth connecting this dialogue with *The Republic*. For *Timaeus*, the root of ignorance lies in a misalignment between the body and the mind. One possible measure to prevent this type of ailment of the soul might be the birth control policy proposed in *The Republic,* as it aims to ensure that great minds are not housed in flawed or inadequate bodies. By regulating reproduction and fostering the ideal combination of physical and intellectual traits, Plato suggests that society could maintain a harmonious balance, minimizing the potential for discord between the body and the mind, and thus promoting the cultivation of wisdom and virtue.

In this article, we have attempted to show a possible tension between the *Timaeus* and the rest of Plato's works concerning moral responsibility, inasmuch as the *Timaeus* seems to endorse a determinism that would contradict the firm resolution found in other dialogues, according to which the moral agent is responsible for their actions. To address this issue, we have referred to Taylor's commentary on the *Timaeus*, and, to explore a potential answer, Cornford's commentary. In this way, we have tried to show how interpreting the words of the characters in Plato's dialogues through their narrative elements—specifically, how the author characterizes them, the role they occupy in the city, or the profession they pursue—can be insightful. Thus, *Timaeus* would not be speaking about human nature as determined, but as determinable by political institutions, interpreted in this manner due to the statesman role Plato assigns to the character. In conclusion, the *Timaeus* would not be in contradiction with the *Republic*, as Taylor suggests, but rather in harmony with it; more specifically, the political-pedagogical project of the *Republic*, which includes population control and regulation of reproductive relationships, would be completed in the medical philosophy of the *Timaeus*: failing to adhere to the legislation of the *Republic* would lead to the soul's diseases described in the *Timaeus*. Therefore, it presents itself as a possible link between the two dialogues.

Bibliographic references

Cornford, F. M. (1997). *Plato's Cosmology,* Routledge & Kegan. (Original work published 1937)

Gill, C. (2000). The Body's Fault?: Plato's Timaeus on Psychic Illness. In M. R. Wright (Ed.), *Reason and Necessity: Essays on Plato's Timaeus* (pp. 59–84). Classical Press of Wales.

Hill, A. (2016). Forms as Active Causes in Plato's Phaedo and Timaeus. In CU Dissertations - Open Access. The Catholic University of America; Dissertations and Theses. Catholic University of America.

Jorgenson, C. (2021). Responsibility, Causality, and Will in the Timaeus. In C. Jorgenson, F. Karfík, & Š. Špinka (Eds.), *Plato's Timaeus, Proceedings of the Tenth Symposium Platonicum Pragen*se (pp. 259–274). Brill.

Kamtekar, R. (2016). The Soul's (After-) Life. Ancient Philosophy, 36 (1):115-132.

Pears, C. D. (2015). Congruency and evil in Plato's Timaeus, *The Review of Metaphysics*, 69(1), 93–113.

Plato. (1968). *The Republic of Plato* (A. Bloom, Trans.). Basic Books.

Plato. (1997). *Plato: Complete Works* (J. M. Cooper, Ed.). Hackett.

Plato. (2000). *Timaeus* (D. J. Zeyl, Trans.). Hackett.

Sedley, D. (2019). The Timaeus as Vehicle for Platonic Doctrine, in Victor Caston (ed.), *Oxford Studies in Ancient Philosophy*, Volume 56, Oxford Studies in Ancient Philosophy (Oxford, 2019; online edn, Oxford Academic, 19 Dec. 2019)

Stalley, R. F. (1996). Punishment and the Physiology of the Timaeus. *The Classical Quarterly*, 46(2), 357–370. http://www.jstor.org/stable/639792

Steel, C. (2001). The Moral Purpose of the Human Body. A Reading of Timaeus 69-72. *Phronesis*, 46(2), 105–128.

Taylor, A. E., (1927). *A Commentary on Plato's Timaeus*. Clarendon press.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 41-58

Artículo/Article

Basically intelligent: ontological and rationalistic perspectives on artificial intelligence in the context of basic income discourse

Básicamente inteligente: perspectivas ontológicas y racionalistas sobre la inteligencia artificial en el contexto del discurso de la renta básica

Shawn Christopher Vigil

Charles University, Czech Republic
2quillswriting@gmail.com
http://orcid.org/0009-0003-8918-9654

DOI https://doi.org/10.48204/2805-1815.8472

INFORMACIÓN DEL ARTÍCULO

Recibido el: 28/5/2025

Aceptado el: 31/8/2025

Keywords:

Applied ethics, basic income, artificial intelligence, political philosophy, ontology

Palabras clave:

Ética aplicada, renta básica, inteligencia artificial, filosofía política, ontología

ABSTRACT/RESUMEN

Abstract:

Basic income is a novel social welfare policy proposal that looks to preserve liberal-egalitarian principles by offering a cash entitlement delivered regularly to every individual in a given society without any stipulations (e.g., work or income requirements). The interest in such kinds of programs has grown larger in the context of exponential technological advancement, with anxieties about the prospect of AI displacing large portions of human labour abounding. However, while the problem of automation has been addressed in the basic income literature, very little philosophical treatment of it has been offered. The present essay aims to fill this gap by elucidating, evaluating, and articulating philosophical arguments that lie at the intersection of AI and ethics. The first argument deals with the question of ontology, viz., whether it is possible in principle for AI to perform all tasks associated with human labour. This argument is explored through a critique of Searle's well-known arguments against the computational theory of mind, together with Dreyfus's phenomenological perspective on the significance of context for sense-making. It is suggested that even if AI might not be able to authentically instantiate intelligence of a general kind, it might nevertheless be capable of adequately performing all tasks associated with human labour. The second argument deals with economic reasoning, viz., whether it would be rational for firms to substitute human labour for AI. It is suggested that micro- and macro-economic rationales betray each other and therefore cannot reliably discount the possibility of significant or complete displacement of human labour. Given that AI remains in principle a possible threat to socio-economic welfare via its relation to labour markets, we end by considering how basic income is uniquely situated to remedy the situation.

Resumen:

La renta básica es una novedosa propuesta de política de bienestar social que busca preservar los principios liberales e igualitarios al ofrecer un derecho a una prestación económica que se entrega regularmente a cada individuo de una sociedad determinada, sin ninguna condición (por ejemplo, requisitos de trabajo o ingresos). El interés en este tipo de programas ha crecido en el contexto del avance tecnológico exponencial, con la creciente inquietud ante la posibilidad de que la IA desplace gran parte del trabajo humano. Sin embargo, si bien el problema de la automatización se ha abordado en la literatura sobre la renta básica, se le ha ofrecido muy poco tratamiento filosófico. El presente ensayo pretende llenar este vacío elucidando, evaluando y articulando argumentos filosóficos que se encuentran en la intersección de la IA y la ética. El primer argumento aborda la cuestión de la ontología, es decir, si es posible, en principio, que la IA realice todas las tareas asociadas con el trabajo humano. Este argumento se explora mediante una crítica de los conocidos argumentos de Searle contra la teoría computacional de la mente, junto con la perspectiva fenomenológica de Dreyfus sobre la importancia del contexto para la construcción de sentido. Se sugiere que, aunque la IA no sea capaz de instanciar auténticamente inteligencia de tipo general, podría ser capaz de realizar adecuadamente todas las tareas asociadas con el trabajo humano. El segundo argumento aborda el razonamiento económico, es decir, si sería racional que las empresas sustituyeran el trabajo humano por la IA. Se sugiere que las lógicas micro y macroeconómicas se contradicen entre sí y, por lo tanto, no pueden descartar con fiabilidad la posibilidad de un desplazamiento significativo o completo del trabajo humano. Dado que la IA sigue siendo, en principio, una posible amenaza para el bienestar socioeconómico a través de su relación con los mercados laborales, concluimos considerando cómo la renta básica está en una posición privilegiada para remediar la situación.

Introduction

It is not as though we have not encountered this narrative before: A benevolent (or at least well- intentioned) intelligence bringing forth its progeny, presumably using itself as the schematic for its design. And this creation is destined to cultivate the world so as to transform it into a paradise that wants for nothing. In myths we have seen this, but nothing quite like it in reality – until now. The first three waves of industrialization rapidly and radically transformed not just the material world and society, but also our understanding of and relationship with them and, perhaps more importantly still, ourselves. The so-called fourth wave in which we are currently enveloped stands to be just as or perhaps even more rapid and transformative. And for the first time in history, the human mind is meeting in the world what it has only previously met in imagination: An intelligence not dissimilar to its own, and even something more.

Artificial intelligence (AI) confronts humanity with deep ontological and practical questions. In regard to the former, it challenges notions of intelligence, consciousness, and humanity as such; in regard to the latter, it forces us to reckon with the possibility that any beings sophisticated enough to perform most, or all human functions will render us redundant, placing us in a precarious socio- economic situation. Basic income discourse occupies itself with questions of the second kind, although, of course, it is underpinned by questions of the first kind. In any register we must ask, how can we ensure that if and when labour becomes exceedingly scarce or disappears altogether, welfare does not vanish along with it? Advocates argue that basic income is the only policy which can provide the security needed in the context of such a society. Others deny the eventuality wholesale, finding nothing especially novel in the most recent wave of technological development (LSE, 2025). If it is true that AI cannot and will not have the radically displacing effects we imagine, then it becomes a non-issue; the argument is irrelevant in any discussion of social welfare, and we can safely leave off with fantastical projections of a post-work society and concentrate on more familiar and realistic arguments. If, however, it would in principle be possible for AI to perform all tasks relevant to human labour – from the most primitive to the most intellectually demanding – then we must seriously consider what safeguards we should have on standby in the case of our eventual complete substitution. The question then becomes, can and will AI threaten human labour such that we are left in a desperate situation that only a basic income can remedy?

Technological innovation is nothing new, nor are its effects on markets and economies. Those who are unconcerned about AI often appeal to history: Economist Heidi Schierholz, for example, observes that when new technologies are introduced, there is indeed temporary displacement in certain sectors, but they are counterbalanced by developments in others, resulting in relative stasis at a minimum and economic growth at best (Vox, 2017). Contrary to the predictions of dystopian alarmists, AI rather seems poised to facilitate general welfare in the form of increased innovation and productivity. The Future of Jobs Report 2025, published by the World Economic Forum (2025), casts AI as a major influence in employment trends in its projection of a net growth of 78 million jobs. And amid fears that as high as 47% of jobs could face technological replacement, more modest calculations can put that number as low as 9%, far less cause for any serious concern (Frey & Osbourne, 2017, p.114; Arntz & Zierahn, 2016). Technological unemployment (that is, unemployment instigated by technological progress) is therefore not near the existential threat that it is sometimes sensationalized to be, as human labour will likely continue to be complemented rather than substituted by automation.

These kinds of observations and arguments provide little comfort for the less optimistically- minded. We can grant that historical trends reveal predictable patterns and nevertheless retain the suspicion that something unprecedented is couched in this new frontier (Ford, 2015). After all, tasks that were replaced in earlier eras were largely mechanical and routine, whereas the capabilities of newer technologies are becoming increasingly sophisticated, their potential seemingly limitless. It is one thing for an automaton to perform the isolated task of assembling specific raw materials at a station in an assembly line, and quite another for it to diagnose skin conditions, evaluate legal documents, produce art, provide advice on personal affairs, write computer code and academic essays, or balance financial accounts. With such promise and uncertainty, we might temper our historically-informed confidence that things will carry on as they always have. Furthermore, the kind of work people will be compelled to pursue as a consequence of technological replacement and unemployment might not be sufficiently accommodating. For example, perhaps workers forced out of their industries simply do not have the interest or talents necessary to adapt to any newly developed sectors; parallel to the previous example of the automaton, a manual labourer could probably as easily chop timber as weld metal, but it would be a perhaps too demanding and even unreasonable expectation that he or she leave such work altogether and learn to code instead. And where would wayward labourers go if the newly developed sectors became unsustainably saturated? These and related possibilities raise further concerns that pernicious features of current economic systems (e.g., inequality) could be exacerbated with technological progress. The 'godfather of AI' and Nobel laureate, Geoffrey Hinton, powerfully characterizes the situation thusly:

> We are talking about having a huge increase in productivity, so there is going to be more goods and services for everybody, so everybody ought to be better off. But actually, it is going to be the other way around, and it is because we live in a capitalist society. And so, what is going to happen is this huge increase in productivity is going to make much more money for the big companies and the rich, and it is going to increase the gap between the rich and the people who are going to lose their jobs… If the profits just go to the rich, that is just going to make society worse. (Nobel Prize, 2024)

A radical shift in policy – and even in institutional structures – could well be in order. It is true that we have encountered technological innovation before, but perhaps nothing quite like this. And even if we are able to adapt to some extent, we might not be able to adapt as we have in the past.

Of course, our predictions are going to vary with our assumptions and methodological choices.

If, according to our preferred methods and observations, we determine that an eventuality is highly unlikely, we might justifiably judge that allocating resources in anticipation thereof would be inefficient and a fortiori unethical, insofar as those resources could have been invested elsewhere and manifestly increased welfare. Some eventualities might, however, be of grave enough consequence that, if we cannot disqualify them outright, we ought to nevertheless have a contingency plan in the ready. To truly allay our concerns, the more effective strategy would be to find principled reasons why automation could not possibly result in the state of affairs that dystopian alarmists imagine. Two arguments readily present themselves: The first is an ontological claim to the effect that AI simply cannot perform some of the important tasks associated with human labour, and the second is an economic claim to the effect that even if such technology could be achieved, it would be irrational to implement it in such a way that significantly displaces human labour.

Understanding Ontology

In order to evaluate the first claim, we must first determine which tasks, if any, associated with human labour could not possibly be done in principle by AI. To date, it is already manifestly evident that many mechanical tasks can be automated, and the list of more intellectually demanding, call them 'cognitive tasks', grows year in and year out. All of these tasks fall under the category of what we call 'weak AI', which is the kind of artificial intelligence capable of performing very well-defined tasks with at least some degree of human oversight. This is contrasted with the notion of 'strong AI', otherwise called 'Artificial General Intelligence' (AGI), which is the kind of artificial intelligence that would be virtually identical to human intelligence. The question becomes, which tasks, if any, associated with human labour require intelligence of this second kind?

Now, when we think about what labour entails, we can deconstruct any given occupation into sets of tasks and skills, where the former are understood as that to be done and the latter as those competencies needed to perform tasks. Take, for example, caretaking: Caretakers must be able to: maintain records, which requires literacy skills (both traditional and digital); assist with domestic chores like cleaning, shopping, or facilitating health regimens, which require physical skills and sometimes special technical skills (like operating automobiles or other instruments relevant to specific industries); communicate with dependents, which requires soft skills (and which, in addition to linguistic skills, also require emotional intelligence, empathy, sound judgment, etc.). Often, caretakers can assume even more demanding roles, such as being moral educators, confidants, or companions.

Thus, to be a competent caretaker is to be able to engage in a variety of tasks using a diverse set of skills in creative ways. As this example clearly illustrates, labour is an incredibly complex phenomenon. But how much of it necessarily evades the potential of AI? That is, how much of this cannot be done in the absence of intelligence of a general kind?

Granted that at this admittedly nascent stage of technological development an entirely integrated machine has not been realized, it takes no great effort of imagination to conceive of a multifunctional automaton. All the ingredients are already there: The caring professions have already begun deploying robot assistants for everything from executing precise surgical procedures, to running errands, managing records, maintaining clean environments, and even providing emotional support (Falcone, 2024; Yazar, 2025). Prima facie, the challenge appears to be merely the technical one of putting everything together. Nevertheless, even if all the requisite skills could be consolidated into a single machine, we might find it wanting in important ways. To be sure, such a machine might be able to perform mechanical and cognitive tasks – sometimes even better than its human counterparts – but it would not be able to do them in the way a human does. While this might not be an issue for some tasks, and for some it is indubitably an advantage, for the most important, uniquely human activities, it might be an insurmountable shortcoming.

The thing that is ostensibly missing, and a fortiori cannot possibly be instantiated in a machine, is authentic understanding. The famous 'Chinese Room' thought experiment formulated by Searle (1980) challenges the computational theory of mind upon which early AI was predicated and seeks to advance the thesis that syntax alone is not sufficient for semantics, or to put it otherwise, that we cannot move from purely formal symbols and operations to meanings. As the argument goes, suppose a monolingual English speaker is isolated in a closed room, equipped with nothing more than a set of materials which instructs how to manipulate symbols so as to produce coherent sentences in Chinese.

Messages in Chinese are anonymously delivered to the individual from outside the room through a small slot, and the individual follows the instruction materials, producing coherent responses and sending them back. To those on the outside, it appears that their interlocutor is a competent Chinese speaker, but in fact he is not; he is simply taking input, manipulating symbols by following a set of instructions, and then generating output. If this is indeed analogous to how computers operate, then they only have the appearance of understanding, rather than authentic understanding. For they no more understand the inputs and outputs than the hypothetical individual in the room understands Chinese.

And since human beings do have authentic understanding, the computational model must be false or otherwise incomplete. Human faculties consist of more than mere computation.

This thought experiment has generated vigorous debate that carries on to this day. Some maintain that while the individual producing the responses according to the script might not have understanding, the system as a whole nevertheless can be said to (cf. Copeland, 2002); others contend that if understanding cannot be attributed to the individual or the system, then it cannot be attributed to human agents either (or else if it can be attributed to one, it should likewise be attributed to the other) (cf. Dennett, 2013). For our purposes, the main question is, what is understanding's role in labour? Are there any tasks that an automaton could not adequately perform without understanding in some deep sense?

Before we answer such questions, we must first elucidate what exactly is meant by understanding. As an intuitive starting point, we might simply claim that it is reasonable to attribute understanding to an agent as long as it demonstrates behaviours associated therewith: If one is given a command and carries out the task appropriately; if one is asked a question and produces a plausible response; if one can pose a relevant question on a topic; etc., then it would seem that for all intents and purposes, and as far as we can possibly know, the agent understands. This is essentially the idea behind the Turing Test in all its iterations. If an automaton can interact as well as any human agent in the environment, then what exactly marks the insuperable difference? If it is something radically subjective (something that 'it is like' to be the thing in question, or a 'beetle' in a box to which only one has access), we might doubt our ability to determine the presence of understanding from without at all, whether in a human or a machine (Nagel, 1974; Wittgenstein, 2009). And if understanding amounts to overt behavioural demonstrations, then what before seemed to be merely appearance becomes less so.

Therefore, we would need a rather different conception of understanding in order to maintain a categorical difference between human and artificial intelligence. Dreyfus (1979), channelling Heidegger and Merleau-Ponty, provides just such an account: What a supposedly intelligent machine lacks is what we might call situatedness. Human beings necessarily find themselves always already in contexts, through which, and only thorough which, the world makes sense. When we understand things, it is not a matter of assembling impersonal, disparate bits of data, analysing them according to prefigured rules, and then operating accordingly, but seeing things as they are to us, encountering them in ways that are conditioned by idiosyncratic histories and interpretations (which inform and work upon each other), and making uncertain but hopeful choices:

> Our sense of the situation we are in determines how we interpret things, what significance we place on the facts, and even what counts as facts for us at any given time. But our sense of the situation we are in is not just our belief in a set of facts, nor is it a product of independent facts or context-free features of our environment… We never get into a situation from outside any situation whatsoever, nor do we do so by means of context-free data. (Dreyfus, 1989, pp. 43-44)

Locating oneself in a personal and global narrative, evaluating and feeling certain ways about the characters and events that populate it, choosing to attend certain of them more or less or rather than others – all of which mutually determine the ways the experiences unfold and continue ever unfolding in a sprawling hermeneutic circle – this is the situation of the human being. And this, presumably, is precisely what the machine lacks. A fully integrated AI might be able to recount all the existing philosophical and scientific scholarship ever recorded, but could it take an interest in any of it? Could it find itself at stake in anything? Could it pose itself an original question that would prompt innovation and change the way it relates to what it pretends to know and shape what it might want to know it the exploration of unfamiliar terrain? The questions that it makes sense for us to ask and the projects to which we choose to dedicate ourselves emerge out of a background of meaning that is not of a purely formal nature. And in the absence of this situatedness, an agent remains suspended in a vacuum, as it were, paralyzed from the lack of sense needed to inspire it to move in a particular direction (and in that particular direction rather than another). “A glaze cleansed of everything past does not see things as they truly are; it sees precisely – nothing” (Reid, 2019, p. 46). To understand is to have a sense of situations, to contemplate and be engaged with this complicated and interconnected world as it discloses itself to us through time.

But how did we find ourselves in this situation, and is it really impossible for a machine to be situated? After all, there was a time before homo sapiens, and a time before collective and individual stories began to be written. Likewise, AI has a history following the progression from mechanization to digitization, and it could also presumably act as if it had a history. In this respect, the parameters appear similar for biological and artificial intelligences. A human agent has and acts as if she has a history; an automaton has and can act as if it has a history. If there is a meaningful difference, we must probe the qualification implied in the modal verb ‘can’. Whereas a human agent in fact has a biography which she also acts as if she has, any biography that an AI would be programmed to act as if it had would amount to a fabrication. The human agent in fact had a mother, experienced the joy of success and the disappointment of failure, nursed the wounds of a broken heart, impacted others, forged enriching friendships. And all these experiences inform and influence the way that she acts in the world that she encounters in every new situation. Meanwhile, the machine might be made to act as if it had similar experiences while in fact having none. Probing such a machine, it could no doubt recount a persuasively rich narrative of historical development, childhood memories, thwarted intentions, and future hopes. And all of it would be as artificial as the intelligence itself. Something about the unreality of these experiences might dispose us to reject the ontological claim that AI can genuinely find itself situated and have understanding in the same deep sense that a human agent is and has. However, we might consider analogous cases in human agents in which we might be hesitant to discount unreal experiences. For example, the ‘alters’ of those suffering dissociative identity disorder (DID) or persons in the throes of dissociative fugues do not in fact have the biographies they recount and feel as though they actually do; in these conditions, they display all the other complex faculties and capacities of ‘real’ persons – they believe certain things to have happened to them, have impressions of states of affairs, and evaluate and feel ways about things as consequences thereof, i.e., they behave as situated and understanding agents. The unreality of their experiences would not permit us to treat them as if they were not worth acknowledging, or worse still, imply that those identities do not remain morally considerable beings. If we want to deny them legitimacy on the basis of things actually having been the case, then we would need to demonstrate how that, and only that, is determinant of authentic situatedness and understanding, rather than the confluence of everything else associated therewith absent actual experience. On the other hand, if acting as if one was situated and understands is sufficient, then the categorical difference between human and machine again begins to fissure.

What matters for this discussion is whether it would be enough for an AI acting as if it was a situated, understanding being to adequately perform activities for which we think situatedness and understanding are necessary. We turn back to the example of the caring professions. Of the tasks with which caretakers are charged, perhaps the most challenging are those related to interpersonal interactions (i.e., those requiring soft skills). Breaking bad news to loved ones, offering emotional support or moral tuition, earning the trust of others, etc. are all highly delicate matters that we might regard as quintessentially human (i.e., tasks associated with intelligence of the second kind). In order to successfully navigate these sorts of situations, one must have a sense of them in the robust way heretofore elaborated. Empathy is comforting because we know that she who empathizes understands our experience, and not just in a descriptive way. We are receptive to advice because she who offers it is someone we trust, who has relevant expertise not just from erudition but also through lived experience. If we were to receive the same kind of support from an AI that has merely been programmed to behave as if it had lived experience to which it could appeal in order to offer empathy and advice, we might not be so receptive and comforted. It might strike us as fraudulent. But then, we also routinely connect with fictions. Are the lessons we take from Antigone and Hamlet less substantive because their ontological status is contentious? Or are they situated, understanding beings only as long as the covers of their tomes remain open? And if they are not, would their appearance as such for the duration of their stories not still have a lasting effect? If we take them to be ‘unreal’ or only real for the duration of their stories, and what we learn from them nevertheless impresses itself upon us lastingly, then why should we discredit an AI whose situatedness and understanding amounted to a fiction?

None of this is to say that there is no difference between something actually having been the case and something only imagined having been the case. If a lover acted as though she had betrayed her beloved, and the beloved, through her living as though that was the case, believed himself to have been betrayed, even though the betrayal never actually occurred, then we might say that they are both living under an illusion, even while pragmatically the betrayal is as real as if it had actually happened. We would consider them obstinate at a minimum if after having learned that the betrayal was an illusion they insisted on carrying on believing it. The point, however, is not whether or not they are mistaken, but only that they have a sense of a situation, whatever it happens to be.

Herewith the ontological gap might not yet be closed. It is not clear that acting as if one is a situated, understanding being is equivalent to being a situated, understanding being. However, it might be enough to convincingly display situatedness and understanding in order to perform those tasks in which they are required. AI can already perform mechanical and cognitive tasks, and it seems possible in principle that it could even perform the most human of tasks. As long as this remains an open possibility, we cannot rest assured that human labour is unquestionably secured.

Economic (Ir)rationality

We can now evaluate the second argument, viz., that even if it is possible to develop AI such that it could perform all tasks relevant to human labour, it would be irrational to implement it to such an extent that it would significantly displace it. According to standard economic models, rational agents act so as to maximize utility. This principle offers a fairly straightforward protocol for firms considering automation: A task should be automated if doing so would result in lower input costs compared to human labour (since lowering input costs would effectively translate to higher profits, i.e., more utility). It follows, therefore, that if automating a significant portion of tasks would be more costly than hiring human labour, then it would be irrational to invest in automation. But would this case ever obtain, and if so, should we expect it to endure?

The premise upon which this argument depends is the empirical one, viz., that either or both the initial investment in or maintenance of labour-saving technology in the form of AI would in fact amount to greater costs for a firm than human labour. Estimates of what it would cost to achieve and deploy AGI (or something sufficiently comparable) at scale are notoriously dubious and speculative at best. But in any case, in order to advance the argument on these grounds, we would need to reject the premise that economic optimists offered earlier, i.e., the historical claim that there is nothing unprecedented in the trends. Of course, new technologies are initially quite expensive, but as they develop, their costs tend to go down. Automobiles and personal computers, once luxuries reserved for the incredibly wealthy, are now ubiquitous. If we are convinced that past trends will continue, then it would seem we should have no reason to suppose that the projected costs of AI will pose a serious obstacle, at least not in perpetuity. As AI becomes more integrated into our economy and society, we should expect costs to become less prohibitive, as has always been the case with novel commodities. To reject this would be to open the possibility that something economically unprecedented is couched in this new frontier, which the economic optimist wants to simultaneously reject, rendering the position inconsistent. It is also worth noting that some of the most ambitious AI enthusiasts apparently have no concern about costs anyway. Sam Altman, CEO of OpenAI, one of the leading organizations in this space, expresses the sentiment, "Whether we burn $500 million a year or $5 billion—or $50 billion a year—I don't care, I genuinely don't… As long as we can figure out a way to pay the bills, we're making AGI…" (Hetzner, 2024).

To carry the argument through to its logical conclusion, let us suppose that the technology is both possible and economically expedient. If a firm is to remain rational, then it would be compelled to substitute its human workforce with fully capable AI, since failing to do so would result in lower utility. But in order to determine whether such a policy would actually be rational, we have to consider what the wider effects of significantly displacing human labour would be. The labourers that face technological unemployment are at the same time the consumers to whom firms intend to sell their goods and services; and as long as their finances are a function largely (and in some cases exclusively) of earned income, it is only too obvious that rendering the workforce inert would be to extinguish the purchasing power of a large portion, if not all, of the consumer base. Consequently, firms would have no one to whom to sell their goods and services, their profits would disappear, and both individual and aggregate utility would plummet. Therefore, significantly displacing the human workforce would be irrational.

Have we any reason to suppose that private firms would carry out this sort of holistic calculation? Again, if history is any indication, we might be wary; simply witness the behaviours of monopolistic robber barons and the very need for anti-trust law. If self-defeating self-interest is not an inherent feature of certain economic systems, we would be forgiven if we suspected that it is nevertheless a feature of the psychology of unscrupulous entrepreneurs who appear to have an insatiable appetite for acquisition. As long as rogue economies and governments, along with the actors that remain all too eager to exploit them, cannot be reliably disqualified even by strong normative principles, we need some kind of systemic mechanisms to safeguard against them.

Basic Income as a Remedy to Technological Unemployment

Insofar as significant technological unemployment remains a realistic possibility, how should we respond? It would seem that existing or potential labour-centric apparatuses (e.g., unemployment insurance or a Negative Income Tax) would likely not be able to meet the demands of an increasingly post-work society, since the labour upon which they are predicated is precisely what would be missing. Naturally, then, we would need some program that is not dependent upon labour to function, and the candidate that appears to be uniquely suited for such a situation is basic income. This is because unlike other programs, basic income is unconditional and universal (that is, available to all citizens of a polity without any stipulations for its receipt), and in a society in which work becomes incredibly scarce or disappears altogether, it would not be viable to place work or wealth conditions on the receipt of resources.

We might still question both of these features, however. Granted that economic contributions to social welfare programs might become obsolete in a post-work society, we might nevertheless ponder noneconomic features thereof, e.g., solidarity. The ideas of desert and reciprocity that run deep in the psychology of social beings might be assumed to persist even in a society in which the economy does not depend upon human labour for production and remuneration, challenging the notion of unconditionality in any prospective basic income program. Susskind anticipates that in such a society, the question would become one of contributive justice, rather than distributive justice (Susskind, 2020). That is, while material provisions will be secured, people will still have the intuition that everyone ought to contribute to society in some way in order to have a share in its products. The failure to satisfy these intuitions could undermine solidarity, leading to social fragmentation. A basic income that is unconditional and universal might adequately respond to problems of distribution by securing the material needs of members of society, but it might be inadequate to respond to problems of contribution. Susskind, therefore, proposes a ‘conditional basic income’ (CBI), according to which people would be required to make noneconomic contributions to society in order to receive benefits. “If some people are not able to contribute through the work that they do, then they will be required to do something else for the community instead; if they cannot make an economic contribution, they will be asked to make a noneconomic one in its place” (Susskind, 2020, p. 192). The kinds of noneconomic contributions could be variable, e.g., community education projects or creation of cultural works. The intuition that people should reciprocate and be deserving must be satisfied in order to preserve social solidarity, and this can be done by requiring noneconomic contributions as a condition for the receipt of benefits.

But imposing such a condition seems to imply that should one not fulfil it, then she is essentially condemned to suffer. In a society such as has been imagined, where economic contributions are no longer functional, noneconomic ones exhaust all else that could conceivably be demanded.

Susskind’s proposal is, therefore, not categorically different than existing social welfare programs that compete with an unconditional and universal basic income. Existing conditional programs stipulate that one should be able and willing to work or else be legitimately indisposed (from age or disability); those who are able but unwilling are not entitled to any benefits. Presumably, the same allowances would remain in a post-work society; we would not expect anything from the elderly or lame, but we would still be demanding of the able but unwilling. In a labour-centric society, they are compelled to make economic contributions or suffer. In a post-work society, they are compelled to make noneconomic contributions or suffer. Though the affix changes, the result remains the same. In making distribution contingent upon contribution, the problems of both never abate. Imposing noneconomic contributions when no other contributions could be made would be to satisfy a psychological imperative rather than a moral one, to prioritize an intense but contingent preference over material and ethical necessity.

This is not to say, of course, that the psychological cannot be at once moral. Dissatisfaction, the thwarting of the will, the absence or erosion of special relationships, etc. all inspire moral action and therefore cannot be dismissed as illusory or otherwise insignificant. The point, however, is that the belief that one should not be entitled to the products of society unless she has contributed thereto cannot take priority when the capacity to contribute has been severed. The satisfaction one feels at the principle of reciprocity being unnecessarily fulfilled pales in comparison to the privation to which one necessarily condemns others in order to feel it.

We might also have reservations about commodifying creative and humanitarian, i.e., noneconomic, activities. Engaging in community or cultural work is often considered a virtue in itself and subjecting it to economic interests could pervert its nature. For example, the over justification effect, in which intrinsic motivation is decreased or even extinguished by the introduction of extrinsic incentives, can demotivate people from engaging in these important projects. The value of this kind of work comes from the work itself and the subjective feelings we have about it. Making it a condition for economic benefits threatens the nature and performance thereof, which could lead to the very social fragmentation a CBI is meant to preserve.

Conclusion

A fully technological society in which human labour is completely replaced by AI remains something of a utopian fantasy. Should it ever be the case that the problems of scarcity and distribution were comprehensively resolved by technological innovation, then our economic systems would be radically transformed, as we might not even need to concern ourselves with the threat of privation at all, nor the conventions of exchange necessitated thereby. It is difficult to imagine non-trivial reasons to exclude anyone in a world cultivated by our artificial progeny that wants for nothing. In a near-full technological society, in which resources fall into the hands of the few still gainfully active members of society, redistribution of society's products would have to be a function of something other than the labour to which the disenfranchised many no longer have recourse. Though these circumstances seem like distant and uncertain possibilities, it would be as irresponsible to dismiss them on such grounds as it would be to pass the ecological buck to future generations. The foregoing has attempted to render it plausible that even if AI does not reach the level of true AGI – that is, intelligence on the level of that of humanity, with authentic understanding and a sense of its place in the context of a world and its complex histories – it could well be possible in principle that all relevant tasks great and small associated with human labour can be satisfactorily performed thereby. There may be something unprecedented in these new frontiers. Our rationality can betray us. And if things cannot be resolutely supposed to carry on the same as they always have, then we must be prepared as we have never been.

Bibliographic references

Arntz, M., Gregory, T., & Zierahn U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. *OECD Social, Employment and Migration Working Papers*, 189, 1-35.

Copeland, B. J. (2002). The Chinese room from a logical point of View. In J. Preston & J.M. Bishop (Eds.), *Views into the Chinese room: New essays on Searle and artificial intelligence* (pp. 104- 122). Oxford University Press

Dennett, D.C. (2013). *Intuition pumps and other tools for thinking*. W.W. Norton & Company.

Falcone, S. (2024, March 21). 6 AI robots that are changing healthcare in 2024. Nurse.org. https://nurse.org/articles/nurse-robots

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Basic Books.

Frey, C.D., & Osbourne M.A. (2017). The future of employment: How susceptible are jobs to computerization. *Technological Forecasting and Social Change*, 114, 254-280.

Hetzner, C. (2024, May 3). OpenAI's Sam Altman doesn't care how much AGI will cost: Even if he spends $50 billion a year, some breakthroughs for mankind are priceless. Fortune.com. https://fortune.com/2024/05/03/openai-sam-altman-microsoft-agi-artificial-generalintelligence-costs/

LSE. (2025, March 11). Is AI really taking our jobs? The future of work explained | LSE research [Video]. YouTube. https://www.youtube.com/watch?v=QqfunA6aSS4

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), (1974), 435-450.

Nobel Prize. (2024, December 18). Nobel minds 2024 [Video]. YouTube. https://www.youtube.com/watch?v=1tELIYbO_U8

Reid, J. (2019). *Heidegger's moral ontology*. Cambridge University Press.

Searle, J.R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3, 417-457.

Susskind, D. (2020). *A world without work: Technology, automation, and how we should respond*. Metropolitan Books.

Vox. (2017, November 13). The big debate about the future of work, explained [Video]. Youtube. https://www.youtube.com/watch?v=TUmyygCMMGA

Wittgenstein, L. (2009). *Philosophical Investigations*: 4th Edition. (P.M.S. Hacker & J. Schulte, Eds.). (G.E.M. Anscombe, P.M.S. Hacker, & J. Schulte, Trans.). Blackwell Publishing, Ltd. (1953).

World Economic Forum. (2025, January 7). The future of jobs report 2025. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/

Yazar, K. (2025, March 18). AI Nurse Robots that are Changing Healthcare. *TechTarget*. https://www.techtarget.com/whatis/feature/AI-nurserobots-that-are-changing-healthcare

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 59-77

Artículo/Article

META-PHILOSOPHICAL SKEPTICISM, SELF-DEFEAT AND PRAGMATIC JUSTIFICATION

ESCEPTICISMO METAFILOSÓFICO, AUTODERROTA Y JUSTIFICACIÓN PRAGMÁTICA

Shih-Hao Liu

University of Miami, United States
sxl1385@miami.edu
https://orcid.org/0009-0005-2620-2377

DOI https://doi.org/10.48204/2805-1815.8473

INFORMACIÓN DEL ARTÍCULO

Recibido el: 31/7/2025

Aceptado el: 31/8/2025

Keywords:

Epistemology, meta-philosophical skepticism, self-defeat, pragmaticism, pragmatic justification

Palabras clave:

Epistemología, escepticismo metafilosófico, autoderrota, pragmaticismo, justificación pragmática

ABSTRACT/RESUMEN

Abstract:

Meta-philosophical skepticism goes that we should suspend our beliefs about philosophical claims. Previously, many argued that prevalent disagreements among peer philosophers motivate the skepticism. One immediate anti-skeptical response is that meta-philosophical skepticism is epistemically self-defeating. In brief, meta-philosophical skepticism calls for the suspension of beliefs about premises deployed in arguments for the very position. This makes the skeptical position ultimately call for belief suspension of itself. Many regard the self-defeat worry as a challenge that meta-philosophical skeptics can hardly meet. In this paper, on behalf of the skeptics, I'll argue that it is possible for meta-philosophical skepticism to sidestep the self-defeat worry with the notion of practical justification. I first contrast traditional evidentialist's view about the ethics of belief which states that beliefs can only be justified epistemically with a pragmatist's view that holds beliefs can also be justified practically. To pragmatists, as long as holding a belief facilitates some practical interest like maintaining a flourishing ordinary life, a belief can be pragmatically justified even if evidence an agent possesses is neutral or silent regarding the justification of the very belief. I contend that the justification of premises deployed in meta-philosophical skeptical arguments can also be explained with pragmatist's view. That is, these premises might be epistemically defeated according to meta-philosophical skeptical arguments. However, skeptics are, in a pragmatic sense, still rational to deploy premises since they exhibit some practical values. While self-defeat can be a serious issue in the epistemic domain, it is not so detrimental in the practical domain since many principles can be implicitly self-defeated but still pragmatically justified.

Resumen:

El escepticismo metafilosófico sostiene que debemos suspender nuestras creencias sobre afirmaciones filosóficas. Anteriormente, muchos argumentaron que los desacuerdos prevalentes entre filósofos pares motivan dicho escepticismo. Una respuesta antiescéptica inmediata es que el escepticismo metafilosófico es epistémicamente autoderrotante. En resumen, el escepticismo metafilosófico exige la suspensión de las creencias sobre los presupuestos utilizados en los argumentos a favor de la propia posición. Esto hace que la posición escéptica exija, en última instancia, la suspensión de creencia en sí misma. Muchos consideran que el problema de la autoderrota es un desafío que los escépticos metafilosóficos difícilmente pueden superar. En este artículo, en nombre de los escépticos, argumentaré que es posible que el escepticismo metafilosófico eluda el problema de la autoderrota mediante la noción de justificación práctica. Primero contrasto la visión evidencialista tradicional sobre la ética de la creencia –según la cual las creencias solo pueden justificarse epistémicamente – con la visión del pragmatista, que sostiene que las creencias también pueden justificarse prácticamente. Para los pragmatistas, mientras mantener una creencia favorezca algún interés práctico, como el mantenimiento de una vida ordinaria floreciente, dicha creencia puede estar justificada pragmáticamente, incluso si la evidencia que posee un agente es neutral o silenciosa con respecto a la justificación de esa misma creencia. Sostengo que la justificación de los presupuestos utilizados en los argumentos escépticos metafilosóficos también puede explicarse mediante la perspectiva pragmatista. Es decir, estos presupuestos pueden estar epistémicamente derrotados según los propios argumentos escépticos metafilosóficos. Sin embargo, los escépticos, en un sentido pragmático, todavía son racionales al utilizar dichos presupuestos, ya que exhiben cierto valor práctico. Mientras que la autoderrota puede ser un problema serio en el ámbito epistémico, no lo es tanto en el ámbito práctico, ya que muchos principios pueden ser implícitamente autoderrotados y aun así estar justificados pragmáticamente.

Introduction

According to meta-philosophical skepticism, we philosophers should suspend our judgment about philosophical claims. Previously, considerations from various directions have been cited in support of this position. Some argue that widespread disagreements among peer philosophers call for suspension of our beliefs regarding philosophical claims (Beebee, 2017; Brennan, 2010; Goldberg, 2013; Kornblith, 2013; Licon, 2019; Ribeiro, 2011; Segal, 2024). Others formulate the skeptical argument based on reflections on the history of philosophy, which reveals a recurring pattern of failures among once-dominant theories (Mizrahi, 2014, 2016). Still others contend that our philosophical beliefs are defeated by evidence of the existence of counterevidence that we have not yet be able to conceive or entertain (Ballantyne, 2013, 2015, 2019; Frances, 2016).

A straightforward objection to meta-philosophical skepticism is that the position is self-defeating (Grundmann, 2019; Paár, 2015, 2016). Briefly, the objection holds that arguments for meta-philosophical skepticism require refraining from believing the very premises on which these arguments rely. If this is the case, then skeptical arguments can hardly get off the ground as we can never justifiably infer the intended skeptical conclusion from their premise(s). Many regard this objection a serious yet seemingly unavoidable challenge for skeptics.

In this paper, on behalf of the skeptics, I argue that meta-philosophical skepticism can sidestep the self-defeat objection by appealing to the notion of pragmatic justification. I begin by contrasting the traditional evidentialist's view about the ethics of belief –which holds that beliefs can only be justified epistemically– with the pragmatist view, which maintain that beliefs can also be justified pragmatically. According to pragmatists, as long as holding a belief serves general practical interests, such as sustaining a flourishing ordinary life, the belief can be pragmatically justified even when the evidence possessed by an epistemic agent is insufficient, or even contrary to the belief itself. I contend that the justification of premises deployed in meta-philosophical skeptical arguments can similarly be account for from a pragmatist perspective. That is, while these premises may be epistemically defeated according to the skeptical arguments themselves, skeptics are nonetheless pragmatically still rational in deploying them, given their value in certain practical dimensions of our everyday lives. Whereas self-defeat constitutes a serious problem in the epistemic domain, it is far less damaging in the practical domain.

The plan for the paper run as follows. First, I present a version of argument for meta-philosophical skepticism as a paradigmatic example and the self-defeat objection it faces. I also discuss some skeptics' responses and explain why they fail. Then, in next section, I introduce the notion of pragmatic justification and explain how the notion helps meta-philosophical skepticism to circumvent the self-defeat objection, and finally, I consider and respond to several potential challenges to my argument.

Two clarifications should be made before proceeding. First, the meta-philosophical skepticism discussed in this paper refers to a *global* version of the position. It targets all (or nearly all) philosophical claims. By contrast, some theorists defend a more *local* form of meta-philosophical skepticism, which specifically challenge particular philosophical methods, such as intuition (Machery, 2017; Alexander, 2012) or inference to the best explanation (IBE) used in metaphysics particularly (Bryant, 2020; Bueno & Shalkowski, 2020). These local versions will not be the focus of this paper. Second, in previous works, many have pointed out that the strategies in arguments for meta-philosophical skepticism can also be used to cast doubt on controversial claims in other domains, such as politics or religion. While I acknowledge this generalizing implication, I will set aside further discussion of it due to space constraints.

Meta-philosophical Skepticism and the Self-defeat Objection

As noted in the previous section, there are various forms of meta-philosophical skepticism, each grounded in different considerations. For the sake of a more focused discussion, I'll present a version of the skeptical argument based on peer disagreements as a toy model for the subsequent discussion. Let *p* represent a given philosophical claim or thesis. Consider the following argument:

(P1) There are widespread peer disagreements among philosophers concerning various philosophical matters.

(P2) If there are widespread peer disagreement among philosophers concerning various philosophical matters, then philosophers should suspend judgment about *p*.

(Conclusion) Philosophers should suspend judgment about *p*.

Following the characterization found in previous literature, I understand a *peer disagreement* as a disagreement concerning some subject matter between different parties who possess roughly the same body of evidence and are approximately equal in intellectual and cognitive capacities. And peer philosophers, as we examine their discussions on various philosophical topics, disagree widely. Not only do philosophers disagree on whether a particular philosophical claim *p* is true, but also on a variety of related matters, like whether other philosophical claims interconnected to *p* (say some implication of *p*) are true, or whether arguments, inferences, methods that are used to support (or refutes) *p* are true or not. From this perspective, (P1) should appear *prima facie* plausible. I'll set aside some of the complexities surrounding the truth of (P1). One might question whether philosophers *actually* disagree about every (or nearly every) philosophical issue, a claim that may require empirical support. (Previously, a survey conducted by Bourget & Chalmers (2014) documents philosophers' judgements on 30 substantive issues in philosophy.) I'll avoid further controversies and assume that there is sufficient degree of peer disagreements among philosophers to render (P1) true for the purposes of this discussion.

Now, if philosophers whom I regard as peers reach different judgments than I do, then whether I'm justified in believing the philosophical claims in question becomes doubtful, as expressed by (P2). Brennan conveys a similar idea, stating that "radical dissensus shows that philosophical methods are imprecise and inaccurate" (2010, p.3). Likewise, Goldberg argues that disagreements among peer philosophers provide defeaters "by way of making salient the possibility that at least one of the disputing parties to the debate is unreliable" (2013, p. 170). Thus, the consideration of widespread disagreement gives rise to an epistemic obligation to suspend our judgement regarding philosophical claims.

A straightforward yet powerful anti-skeptic objection to the argument above is that it is *self-defeating*. For an argument to be persuasive, it seems minimally true that we should be epistemically justified to believe or know its premise(s). However, according to the argument itself, we should refrain from believing in the premise(s) deployed in it. Consider substituting (P2) for *p* in the argument above. If it is plausible to expect a widespread disagreement among philosophers regarding the truth of (P2), then we should find whatever epistemic justification of (P2) we have being defeated by peer disagreements as suggested by the argument itself and hence should refrain from believing in (P2). A similar line of reasoning applies to (P1). Grundmann (2019) argues that the plausibility of (P1) depends on our ability to identify epistemic peers, at minimum, by evaluating their track records. However, such identification is possible "only if one presupposes that those philosophical beliefs that form the basis of track-record evaluations are justified" (p. 224). Yet, the skeptical argument itself instructs us to suspend judgment about those very beliefs. As a result, (P1) is also defeated. If we are not justified in believing either (P1) or (P2), then we have no reason to accept the argument's reasoning or endorse its skeptical conclusion. In this way, the skeptical argument fails to get off the ground.

This line of self-defeat objection can be further generalized to other versions of skeptical arguments for meta-philosophical skepticism. Consider, for instance, a version of the skeptical argument grounded in a form of pessimistic historical meta-induction:

(P1*) The history of philosophy reveals a pattern in which past philosophers failed to entertain or conceive serious objections to what were then considered the most promising philosophical theses.

(P2*) Present day philosophers likewise fail to entertain or conceive serious objections to what are now considered the most promising philosophical theses, including *p*.

(P3*) If (P2*) is true, then we should suspend our judgment about *p*.

(Conclusion) We should suspend our judgement about *p*.

In a series of examples cited by Mizrahi, several past accounts that were once regarded as the most promising ones were later found to face serious objections that their original proponents failed to conceive. For instance, the traditional justified true belief (JTB) analysis of knowledge was challenged by Gettier (1963). Or the descriptive theory of proper name was later met with Kripke's influential modal objection (Kripke, 1980). These ample examples in the history show a pattern which render (P1*) to be true. Of course, it is controversial whether this historical pattern has been exaggerated. Nevertheless, I'll take (P1*) to be at least *prima facie* plausible and proceed accordingly. Now, if we assume the pattern described in (P1*) holds, we can inductively infer (P2*). For any given philosophical thesis that we're currently considering and even provide the defenses with our best effort, we have reason to expect that there are (or will be) unconceived serious objections which we are currently not able to address. As Mizrahi puts it:

> …*the history of philosophical inquiry offers a straightforward rationale for thinking that there typically are serious objections to our best philosophical theories, even when we are unable to conceive of them at the time*. (Mizrahi, 2014, p. 426, his italics)

If we admit that this is the case, then a straightforward response, as suggested by (P3*), is to suspend our judgement about the given philosophical thesis.

However, from the perspective of anti-skeptics, the inductive argument above is also self-defeating. Consider (P3*). According to the argument's own logic, we should expect unconceived serious objections to (P3*) and thus ought to suspend judgement regarding it. But this defeats the very inductive argument for meta-philosophical skepticism, as we are no longer justified in believing one of its core premises. Anti-skeptics argue that this objection can be generalized to all versions of arguments for meta-philosophical skepticism. Regardless of which epistemic principles skeptics invoke, if those arguments also prescribe suspending belief in those very principles, then the arguments fail to pose any real threat to our knowledge or justification with respect to philosophical claims.

Skeptics are aware of the problem and have offer their responses. Skeptics might try to argue against the self-defeat objection by stating that their skeptical arguments work in a *parasitic* manner. That is, skeptics themselves need not endorse the premises of their arguments. Rather, as long as the readers of those arguments are committed to the relevant premises, the force of the skeptical conclusion remains intact. I find this line of response unpersuasive. If the self-defeat objection reveals that skeptical arguments defeat themselves, then the rational response for readers is simply to refrain from believing those premises –regardless of any prior commitment they may have had before encountering the skeptical argument.

Another line of response holds that it is possible that the justification of premises in skeptical arguments is secured from being compromised by arguments themselves since the premises happen to be few justified philosophical claims. For instance, Brennan states that:

> However, it may just be that a small set of philosophical issues is answered and that philosophical issues is answered and that philosophical methodology works reliably on a small set of issues, i.e., just in the areas needed to make the sceptic's argument. For instance, perhaps the sceptic needs probability, an account of the notion of an epistemic peer, some notion of reliability, and not much else. (Brennan, 2010, pp. 8-9)

Brennan's response is hardly convincing. One might reasonably ask what explains the reliability of the methodology invoked by the skeptic's account. If no explanation can be provided, then Brennan's defense is explanatory unsatisfactory and *ad hoc*. On the other hand, if an explanation for the methodology's reliability is offered, then it must appeal to some methods or background account(s). However, if peer disagreement among philosophers is as widespread as (P1) suggests, then we should expect that there are peer disagreements concerning the reliability of the invoked method(s) and background account(s). In this way, the worry of self-defeat resurfaces, as Paár observes: "But if peer disagreement shows unreliability, then surely our method in answering the meta-level question of philosophy's reliability and the epistemic status of our philosophical theories is also unreliable". (2015, p. 32)

Ballantyne (2019) offers yet another line of response by reformulating the skeptical argument with the notion of partial defeater. In contrast to a full defeater which demands us to give up the target belief, a partial defeater only demands us to lower our confidence without fully relinquishing it. According to Ballantyne, it is possible that the skeptical argument defeats its own premise(s) only *partially*:

> If the first-order evidence supporting our belief in the method is strong, then the competence defeaters may push down our confidence only a little. The method may call for some doubt about itself, but not enough doubt to properly eliminate our belief in it. (Ballantyne, 2019, pp. 254-255)

Ballantyne's partial defeater response faces two main problems. First, it weakens the force of arguments for meta-philosophical skepticism. Anti-skeptics can simply dismiss such argument by noting that they merely present *partial* defeaters that demand us to slightly reduce our confidence, while continued beliefs in philosophical claims remain epistemically reasonable. Second, why skeptical arguments merely present partial defeaters instead of full defeaters against their premises needs an explanation. If the explanation can be fully defeated by applying skeptical arguments, then the self-defeat objection creeps back since the epistemic possibility for skeptical arguments to be *full* defeaters against their premises is back in the picture.

Skeptics might attempt to deal with the self-defeat objection by drawing on prior discussion about conciliationism in the epistemology of disagreement, since the view faces a structurally similar self-defeat objection. According to conciliationism, one is rationally required to suspend judgement or reduce confidence in a proposition *p* when confronted with peer disagreement. It is not difficult to see that such a view is self-defeating once we consider the situation where there is a disagreement about whether conciliationism itself is correct or not. Following the prescription of conciliationism, its advocate should suspend or reduce confidence in conciliationism itself. Namely, "your view on disagreement requires you to give up your view on disagreement" (Elga, 2010, p. 179). In what follows, I'll briefly consider two conciliationists' responses to this objection and evaluate whether they help skeptics in replying to the self-defeat challenge.

First, Elga argues that "in order to be consistent, a fundamental policy, rule, or method must be dogmatic with respect to its own correctness" (2010, p. 85). In other words, to maintain consistency, the relevant belief about the correctness of a method is exempt from being defeated by the method itself. Thus, in order to remain consistent, it is epistemically permissible for conciliationists to exempt their very thesis from being defeated by itself. By analogy, skeptics might claim that in order to preserve consistency, premises in skeptical arguments should also be exempted from being defeated by those very arguments.

Second, Pittard (2015) argues that regardless of how the conciliationist response, a commitment to conciliationism is preserved at some level. He illustrates this by distinguishing between belief-credence level and the reasoning level. He contends that if a conciliationist reduces the credence of the belief in the presence of a steadfast opponent, then although the commitment of conciliationism is violated at the belief-credence level by deferring to a steadfast opponent, however, the commitment is still preserved in the reasoning level. On the other hand, if a conciliationist stick to the thesis, then the commitment to conciliationism is violated at a reasoning level, and deference toward a steadfast view is demonstrated. But still, the commitment is preserved at the belief-credence level. Pittard concludes conciliationism is rationally committed to its very idea (although at a different level) either way. Similarly, skeptics can adopt Pittard's strategy and argue that either refraining from believing premises (like (P1) or (P2) above) in skeptical arguments or continuously believing in them preserves the rational commitments to meta-philosophical skepticism.

I contend that both Elga's and Pittard's responses are unhelpful to skeptics here. One immediate follow-up question to skeptics that adopt Elga's line of response is why the preservation of consistency should be valued. It seems that nothing stops us from applying skeptics' arguments to cast doubt on the commitment of consistency preservation as a guidance principle for preservation/rejection of adopting a method or principle. From this perspective, the self-defeat objection re-emerges. And even if there's a way to elucidate why consistency preservation should be respected without being defeated by skeptical arguments, Elga's response is still problematic. There are many ways to preserve consistency when skeptics face the self-defeat objection. One way is to exempt premises in skeptical arguments from being self-defeated. The other is to reject these premises straightforwardly. So *why* must we favor exempting the argument from self-application if there are other moves for the sole consideration of consistency preservation? Framed this way, adopting Elga's strategy is ill-motivated. Adopting Pittard's line of response suffers a similar problem. If neither (1) refraining from believing premises of skeptical arguments according to the reasoning of these arguments themselves, nor (2) sticking with their beliefs about the premises and refusing self-application of the argument violates skeptics' commitment to their skepticism, then what reason is there to favor of (2) over (1)? Skeptics might want to seek for other conciliationists' responses to the charge of self-defeat in hope of resolving skeptics' own problem. I think the hope is dim. In its nature, while conciliationists' responses seek to establish a positive thesis still, however, skeptics seek to argue against all cases of philosophical knowledge and justification. This puts skeptics in a difficult position to uptake conciliationists' strategies without invoking further commitments in some epistemic principles which they should reject.

From the above discussion, we can summarize that these previous attempts to take up the self-defeat objection falls into a trilemma. First, these responses might *weaken* the force of skepticism when considerations like partial vs. full defeaters or consistency preservation are introduced—since philosophical claims that are relevant to these considerations should be granted as possessing certain form of justification. Second, if the invoked considerations can be challenged by skeptical arguments again, then the self-defeat objection returns. Thirdly, if the responses only make space that explain the epistemic permissibility of sticking to beliefs about premises in skeptical arguments along with other viable options, then the responses are again *ill-motivated* and *insufficient* to retain the force of skeptical arguments without favoring of meta-philosophical skepticism non-arbitrarily. In sum, these recent attempts to tackle the self-defeat objection utterly fail.

Is meta-philosophical skepticism doomed due to the self-defeat objection? Not necessarily. In the following section, I'll introduce the notion of pragmatic justification and explain how such a notion can help skeptics to respond to the objection.

Pragmatic Justification in Rescue

Previously, the discussion of the self-defeat objection against meta-philosophical skepticism is heavily revolved around *epistemic* terminologies related to knowledge, justification, reliability, and rationality. Namely, what we should or shouldn't believe solely depends on whether it brings epistemic goods like rationality, understanding, or knowledge. Call such a position of belief evaluation as *evidentialism*. Its core principle is characterized by Feldman as follows: "When adopting (or maintaining) an attitude toward a position, *p*, a person maximizes epistemic value by adopting (or maintaining) a rational attitude toward *p*" (2000, p. 685). If, pace evidentialism, whether we should adopt a belief *only* depends on the expected epistemic values, then indeed the self-defeat objection poses a threat to meta-philosophical skepticism, as we should never rationally reach the intended skeptical conclusion via skeptics' arguments. But it is not always the case that a belief's justificatory status can be solely evaluated with epistemic values.

Recently, a series of works has defended the view that a belief can be justified *pragmatically*, even when the evidence available to the subject is insufficient or neutral to provide epistemic justification for it (McCormick, 2015, 2020; Rinard, 2021). Call this view *pragmatism* in belief evaluation. I shall say more on what pragmatic justification is about below.

There are several notable features of the notion of pragmatic justification. First, in contrast to epistemic-value-related notion like knowledge, truth, epistemic rationality, pragmatic justification is a notion that is more encompassing. According to Rinard, besides epistemic sense of "ought," there is a sense of "ought" that "takes into account all relevant considerations and is in that sense all-things-considered" (Rinard, 2021, p. 441). Sosa adheres and further states that "a belief can be epistemically irrational though rational all things considered" (2010, p. 34). Second, pragmatic justification is *guidance-giving* (Rinard, 2021, p. 441). Namely, it informs us what to do or believe as all-things-considered reasons are balanced. Thirdly, I take that the pragmatic justification, in many cases, can be *partially indifferent* regarding various aspects. A belief can be pragmatically justified even if some (or many) relevant issues regarding its truth have *not yet* been fully settled. This feature in line with many of our ordinary epistemic and linguistic practices as Eklund in his discussion of ontology states that "even genuinely literal assertions have what we may call *non-serious* features, features that are not important to the point of the assertions, and among these features are normally the ontologically committing ones" (2005, p. 558). For example, Eklund make a case with the discussion of ontology of ordinary objects:

> In the case of middle-size dry objects, suppose that the oracle tells us that a radical stuff ontology is correct (there are no objects but only stuff), or that mereological essentialism is correct, or that van Inwagen-style eliminativism (organisms are the only complex objects there are) is correct. In each case, I am as inclined to believe that we would 'go on as before' as I am inclined to believe this in the case of mathematics. Perhaps matters would stand differently if the oracle gave some other type of positive account of why there aren't any middle-sized objects as we conceive them—that it is all a dream or that Berkeleian idealism is correct. But however, that may be, the general point stands. (Eklund, 2005, pp. 559-560)

In ordinary context, we won't take assertion and belief of a subject that there are middle-sized ordinary objects to be irrational even if the subject has not yet believed in a well-established view in ontology or just is completely ignorant about the metaphysical disputes. Fourth, pragmatic justification is overall *consequential*. That is, whether a belief is justified in a pragmatic sense or not depend on the expected outcome of belief possession. But what exactly is the outcome that should be considered? Finally, following McCormick's characterization, I suggest that the outcome should be about our general interests in having a flourishing life. The general interest mentioned here should be distinguished from merely instrumental or prudential interest (2020, p. 8608). While different individuals might have different personal aims and goals in different scenarios or situations, however, some states are, in general, desirable to most individuals. For instance, facilitation of communication, increase in the survival rate, conceptual clarity, etc. (Here, I'm open to the possibility that general interests can depend on both objective (facts) and subjective (like a subject's beliefs) factors.) Summing up, we may understand a belief as being pragmatically justified as long as possessing it is to be expected to fit our general interests that lead to a flourishing life even if it lacks evidential support to a certain extent: "But if some beliefs that are integral to flourishing cannot be grounded in evidence and their truth-value remains indeterminate, this will not detract from their value". (McCormick, 2020, p. 8608)

Some might still think that the possibility of justifying a belief in a pragmatic sense sounds incredible. But there are plenty of cases where pragmatic justification best fits the explanation: "For example, many believe in God despite taking themselves to lack evidence. Or one may be sure that a friend of theirs is innocent, even if they acknowledge that the evidence suggests they won't succeed". (Rinard, 2021, p. 447) Or consider McCormick's discussion on Nozick's take on believing that his children are not automata:

> [Nozick] says even if all the evidence available to him would be the same if his children were automata, so that he cannot know that his children are not automata, this does not undermine his belief that his children are not automata. (McCormick, 2020, p. 8604)

Of course, the cases above like believing god's existence, friend's innocence, or his children not being automata, judging from an evidentialist's perspective, are surely irrational. But this then ignore other aspects where we might still want to claim that these beliefs are somewhat reasonable. And if evidentialist's criterion is the only reasonable one for believe evaluation, then many of laymen's beliefs should probably be charged with irrationality as these beliefs are (from a philosopher's perspective) disappointedly and unsophisticatedly coarse-grained and can be easily defeated with various philosophical arguments. But surely charging that our ordinary beliefs are massively irrational is to a certain extent, undesirable. Adhering to this point, in their works, both McCormick and Rinard appeal to pragmatic justification to explain the rationality of laymen's beliefs about the external world. According to them, even if an individual fully endorses or cannot respond to arguments for external world skepticism, her beliefs about external world object like trees, people, furniture are still pragmatically justified.

Appealing to the notion of pragmatic justification, McCormick contends, can also partially explain why we think epistemic values related to truth, rationality, or knowledge are important to us as they either direct or indirectly contribute to a flourishing life, as she says "by tying epistemic value to the practical, broadly construed, we can make sense of why epistemic norms have the force that they do" (2020, p. 8607). McCormick speculates that the normative force of epistemic reason and justification is ultimately based on pragmatic justification. Echoing McCormick, Rinard expresses some doubts about an autonomous realm of epistemic reason/justification as she says, "my own view is that putative epistemic sense of 'should,' 'reason,' 'justified,' and 'rational' are not in a good standing" (2021, p. 442). I do find both McCormick's and Rinard's view appealing. It seems true that why we care about epistemic reasons is indeed heavily motivated by varieties of practical considerations. However, whether there's an independent realm of epistemic rationality in good standing does not need to be settled here. I'll leave the issue for future research.

So, how does the pragmatic justification help meta-philosophical skeptics to deal with the self-defeat objection? I suggest that skeptics can explain how premises like (P2) in section 2 is upheld even facing the self-defeat charge by stating that we do have pragmatic justification to (P2) or similar principles. Consider the following scenario. Two sources of information are in conflict regarding the opening hour of the local library. Say Sam claims that it should open at 8 a.m. on weekday. But Emma disagrees and states that it should open at 9 a.m. on weekday. Suppose we don't have any reason to discredit either Sam or Emma and there's no pressure or immediate practical consequence if we do not decide which side is correct. It seems that it will be a pragmatically rational move to suspend our judgment for now. Here, we might formulate the principle that guide our consideration as:

(P) Suppose there's no immediate practical consequences for belief suspension; we should suspend our judgement facing disagreement between equally reliable information sources.

Two things to be noted for the formulation of (P). First, the qualifier for "no immediate practical consequences for belief suspension" is intended to resemble the principles under the examination in philosophical activities where there's no immediate harm or other bad outcomes would occur if judgement were not made in time. Second, my use of the term "reliable" should not be understood as the technical epistemic term. What I suggest is that the term is understood in a more non-committal (or partially indifferent) way. That is, in our daily practices, before we have a more refined and stringent philosophical definition about what makes an information source reliable, we already possess some coarse-grained conception of what makes the source reliable (just to slightly formalize with some philosophical precision, think about a disjunction of features that we would associate with reliable source in our daily lives). One will immediately see that (P) is also self-defeating as it is possible that there are equally reliable sources disagree about whether (P) is true. However, even if this is the case, it should be hard to deny that in our daily practices, when we face disagreements, we'll appeal to (P) or other analogous principles to guide us. It is surely that (P) and analogous principles suffer from all kinds of issues if we examine them from a philosophically sophisticated perspective. But it can be hard to deny that at most of the circumstances, following (P) is still a pragmatically rational as it prevents us from the bad consequences with making hasty decisions. In the same vein, we can maintain our beliefs in (P), (P1), (P1*) as even though holding them invoke self-defeat in a more stringent philosophical sense. This is similar to how Rinard argues that individuals (what she refers to as Pragmatic Skeptics) can at the same time appreciate or even be convinced by arguments for external world skepticism but still possess beliefs about the external world as she describes that "Pragmatic Skeptics will exhibit systematic, ongoing diachronic inconsistency in their beliefs" (Rinard, 2021, p. 436). And meta-philosophical skeptics can safely appeal to premises like (P1) as a parcel of their skeptical argument and suspend their judgement on various philosophical theses.

Meta-philosophical skepticism in conjunction with the notion of pragmatic justification. Might still believe or assert claims about knowledge, justification, or reliability. But it should be kept in mind here that skeptics will believe or assert in a more partially indifferent or non-serious sense which does not commit to any philosophically sophisticated sense of truth or epistemic conditions.

While on behalf of skeptics, I argue for the possibility that premise(s) used in arguments for meta-philosophical skepticism can be rational in a pragmatic but *not* in an epistemic sense. I take it that one of the consequences of my defense is that it also retains the possibility for us to believe in some philosophical theses with practical justification.

That is, some philosophical theses can still be believed, albeit we lack sufficient epistemic justification. Like, even if some version of the principle of utility is still under extensive disagreement about its correctness among peer philosophers, under certain circumstances, subjects can still hold it as long as it happens to promote some of the general interests. I suggest that this is not a negative consequence.

Several clarifications should be made here. First, my view should not be conflated with a Moorean commonsense view. Indeed, under my view, a subject can hold many beliefs about commonsense with justification. While what I do commit to is that these beliefs enjoy pragmatic justification as having them contribute to a flourishing life for average human beings, I do not adhere to a Moorean view as such a view still attempt to explain the rationality of our beliefs about commonsense with epistemic values. Also, I take that it is possible for a subject to believe in some claim that is largely in conflict with commonsense as long as we have pragmatic reason(s) to believe in it. Second, Although the notion of pragmatic justification does share some similarity with epistemic notion of justification and knowledge under contextualist's framework (DeRose, 1995; Lewis, 1996), however, it possibly diverges from contextualism regarding, for instance, how notion like general interest should be understood. While contextualist would probably deny, I'm open to the possibility that there are some general interests that contributes to flourishing life in all contexts in an objective sense. Thirdly, my appeal to pragmatic justification is also distinct from Wright's view of epistemic entitlement (2004). According to Wright, it is rational to accept some propositions (what he called "cornerstone propositions") if these propositions are important to some of our cognitive projects. The acceptance here, he contends, should not be understood as beliefs since beliefs respond to evidence. Instead, acceptance should be understood as trust:

> It is in the nature of trust that it gets by with little or no evidence. That is exactly how it contrasts with belief proper, and it is not per se irrational on account of the contrast. Entitlement is rational trust. (Wright, 2004, p. 194)

Wright's view can be criticized from the perspective of a phenomenological consideration. Phenomenologically, both laymen and we don't *just* trust but believe in the existence of external world. This renders Wright's view unsatisfactory regarding the explanatory lacuna regarding how we explain such a belief. Pragmatic justification, on the other hand, scores much better as it both provides explanation and fits with the phenomenological adequacy. The meta-philosophical skepticism with pragmatic justification defended here should also not to be conflated with Pyrrhonian skepticism. Previously, Pyrrhonian skepticism is often charged to lead to an unlivable life since it "counsels' radical suspension of judgment, which could, as Hume suggested lead to a potentially fatal inability to act" (Rinard, 2021, p.440). In contrast, with the notion of pragmatic justification, the rationality of our many ordinary beliefs is explained. In response to the worry that Pyrrhonism leads to an unlivable life, Frede (1980) argues that it is possible for Pyrrhonian skeptics to hold beliefs about "something evident, something that seems to him to be the case" (1980, 194) without commits to the truth of what is believed. Some might worry that this makes my view quite similar to a version of Pyrrhonian skepticism. While I think it can be an interesting issue to further look into, let me just point out there's still a dissimilarity even if Frede's defense is a plausible one. As Frede construe what Pyrrhonian can believe with what's evident, in comparison, I suggest that the range of what meta-philosophical skeptics is wider as it includes what fits general interests. Noted that what fit general interest might not appear to be true or evident. I believe this makes my view more lenient regarding what we can rationally believe and better fit with our ordinary practices.

Conclusion

To summarize, the introduction of pragmatic justification enables meta-philosophical skepticism to sidestep the self-defeat objection. It allows for the possibility that one can be pragmatically rational in believing the premises of skeptical arguments, even if those arguments epistemically defeat those very premises. Pragmatically speaking, we are still justified in holding these beliefs and employing them in argumentation. In this way, the self-defeat objection, at least within the epistemic domain, is deflated. That said, at least two lingering questions merit further discussion. First, is there an independently grounded realm of epistemic evaluation with normative force, or is all normative force ultimately derived from pragmatic considerations? Second, if meta-philosophical skepticism is correct, what role remains for philosophy? Due to limitations of space, I leave these questions for future investigation.

References

Alexander, J. (2012). *Experimental Philosophy: An Introduction*. Polity Press.

Ballantyne, N. (2013). Counterfactual Philosophers. *Philosophy and Phenomenological Research*, *88*(2), 368-387.

Ballantyne, N. (2015). The Significance of Unpossessed Evidence. *Philosophical Quarterly*, *65*(260), 315-335.

Ballantyne, N. (2019). *Knowing Our Limits*. Oxford University Press.

Beebee, H. (2017). Philosophical Scepticism and the Aims of Philosophy. *Proceedings of the Aristotelian Society*, *118*(1), 1-24.

Bourget, D. & Chalmers, D. J. What Do Philosophers Believe? *Philosophical Studies*, *170*(3), 465-500.

Brennan, J. (2010). Scepticism about Philosophy. *Ratio*, *23*(1), 1-16.

Bryant, A. (2020). Keep the Chicken Cooped: The Epistemic Inadequacy of Free-Range Metaphysics. *Synthese*, *197*(5), 1867-1887.

Bueno, O. & Shalkowski, S. A. (2020). Troubles with Theoretical Virtues: Resisting Theoretical Utility Arguments in Metaphysics. *Philosophy and Phenomenological Research*, *101*(2), 456-469.

DeRose, K. (1995). Solving the Skeptical Problem. *Philosophical Review*, *104*(1), 1-52.

Elga, (2010). How to Disagree about How to Disagree? In T. Warfield & R. Feldman (Eds.), *Disagreement* (pp. 175-186). Oxford University Press.

Eklund, M. (2005). Fiction, Indifference, and Ontology. *Philosophy and Phenomenological Research*, *71*(3), 557-579.

Feldman, R. (2000). The Ethics of Belief. *Philosophy and Phenomenological Research*, *60*(3), 667-695.

Frances, B. (2016). Worrisome Skepticism about Philosophy. *Episteme*, *13*(3), 289-303.

Frede, M. (1980). *Essays in Ancient Philosophy*. University of Minnesota Press.

Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, *23*(6), 121-123.

Goldberg, S. (2013). Defending Philosophy in the Face of Systematic Disagreement. In D. E. Machuca (Ed.), *Disagreement and Skepticism* (pp. 277-294). Routledge.

Grundmann, T. (2019). Why Disagreement-Based Skepticism cannot Escape the Challenge of Self-Defeat. *Episteme*, *18*(2), 224-241.

Kornblith, H. (2013). Is Philosophical Knowledge Possible? In D. E. Machuca (Ed.), *Disagreement and Skepticism* (pp. 260-276). Routledge.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.

Lewis, D. K. (1996). Elusive Knowledge. *Australasian Journal of Philosophy*, *74*(4), 549-567.

Licon, J. A. (2019). Modest Meta-philosophical Skepticism. *Ratio*, *32*(2), 93-103.

Machery, E. (2017). *Philosophy Within Its Proper Bounds*. Oxford University Press.

McCormick, M. S. (2015). *Believing Against the Evidence: Agency and the Ethics of Belief*. Routledge.

McCormick, M. S. (2020). Value beyond Truth-Value: A Practical Response to Skepticism. *Synthese*, *198*(9), 8601-8619.

Mizrahi, M. (2014). The Problem of Unconceived Objections. *Argumentation*, *28*(4), 425-436.

Mizrahi, M. (2016). Historical Inductions, Unconceived Alternatives, and Unconceived Objections. *Journal for General Philosophy of Science*, *47*(1), 59-68.

Paár, T. (2015). Disagreement, Self-Refutation and the Minority Report of the Meta-Skeptics. *Kriterion – Journal of Philosophy*, *29*(3), 23-44.

Paár, T. (2016). Meta-Skepticism, Experimentalism, Cartography: The Dead-Ends of Philosophy. In M. Gyöngyösi, Z. Kapelner, Z. Ádám & I. Faragó-Szabó (Eds.), *On What It Is: Perspectives on Metaphilosophy* (pp. 63-85). Philosophy Workshop of Eötvös József Collegium.

Pittard, J. (2015.) Resolute Conciliationism. *Philosophical Quarterly*, 65(260), 442-463.

Ribeiro, B. (2011). Philosophy and Disagreement. *Crítica: Revista Hispanoamericana de Filosofía*, *43*(127), 3-25.

Rinard, S. (2021). Pragmatic Skepticism. *Philosophy and Phenomenological Research*, *104*(2), 434-453.

Segal, A. (2024). Systematicity and Skepticism. *American Philosophical Quarterly*, *64*(1), 1-18.

Sosa, E. (2010). *Knowing Full Well*. Princeton University Press.

Wright, C. (2004). Warrant for Nothing (and Foundations for Free). *Proceedings of the Aristotelian Society*, 104, 167-211.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 78-93

Artículo/Article

Is motor representation a potential answer to the problem of causal deviance?

¿Es la representación motora una posible potencial respuesta al problema de la desviación causal?

Yukun Chen

University of Miami, United States
yukunchen99@gmail.com
http://orcid.org/0009-0006-4200-7181

DOI https://doi.org/10.48204/2805-1815.8474

INFORMACIÓN DEL ARTÍCULO

Recibido el: 6/6/2015

Aceptado el: 31/8/2025

Keywords:

Philosophy of action, intention, action, casual deviance, motor representation

Palabras clave:

Filosofía de la acción, intención, acción, desviación causal, representación motora

ABSTRACT/RESUMEN

Abstract:

This paper examines the problem of causal deviance in theories of intentional action and the role of motor representations in trying to solve it. The problem arises when an agent's intention and action correspond to an outcome, but there is a deviance in the causal chain causing the outcome accidental rather than intentional. Four current theories that involves proposing motor representations as a solution are critically analysed: the Deferral View, Motor Schema View, Dual Content View, and Same Format View. While insightful. I then present a case study involving causal deviance for motor representations themselves, arguing that even with motor representations, the accidental nature of action outcomes cannot be ruled out under a causalist framework. Finally, I suggest moving beyond causalist views, drawing inspiration from an alternative view that cognitions merely bias rather than causally produce motor representations and actions.

Resumen:

Este artículo examina el problema de la desviación causal en las teorías de la acción intencional y el papel de las representaciones motoras en el intento de resolverlo. El problema surge cuando la intención y la acción de un agente corresponden a un resultado, pero hay una desviación en la cadena causal que causa que el resultado sea accidental en lugar de intencional. Se analizan críticamente cuatro teorías actuales que proponen representaciones motoras como una solución: la Visión del Aplazamiento, la Visión del Esquema Motor, la Visión del Contenido Dual y la Visión del Mismo Formato. Si bien esclarecedor. Luego presento un estudio de caso que involucra la desviación causal para las propias representaciones motoras, argumentando que incluso con representaciones motoras, la naturaleza accidental de los resultados de la acción no puede descartarse bajo un marco causalista. Finalmente, sugiero ir más allá de las visiones causalistas, inspirándome en una visión alternativa de que las cogniciones simplemente sesgan en lugar de producir causalmente representaciones y acciones motoras.

The problem of causal deviance

Intentional actions are those executed by agents who own agency over their bodily movements. An essential aspect of elucidating intentional actions is agent's intention, which typically refer to the mental states of committing to action plans, thereby causing the actions. However, explaining actions in terms of intentions is complicated by the problem of causal deviance, a notorious problem proposed by Davidson (1963). It arises in cases where an agent has the intention to act, but there seems to be a deviance in the causal chain leading to the action execution. Davidson's famous example involves a man named Jones who intentionally administers a lethal dose of poison to another man, Smith, with the intention of killing him. However, unknown to Jones, Smith unexpectedly dies from an unrelated heart attack moments before the poison could take effect. In this scenario, the expected cause of death (the poisoning) never occurs. But it still seems intuitive to attribute the death to Jones due to his intentional action of administering the poison with the aim of killing Smith.

In this case, the standard conditions of the causalist views (e.g., Davidson, 1971/2002a, 1978/2002b); Paul, 2009; Bratman, 1984, 1987) for an intentional action are met –Jones' intention to kill, as well as his bodily movements successfully administered the (ineffective) poison, and the outcome of Smith's death. But there is an intuitive sense that their action did not cause the death in the right way for it to truly count as an intentional action. Arguably, this problem shows an explanatory gap between agents' intentions and actions. Many answers have been proposed, but there is ongoing debate about how to best handle such counterexamples within theories of intentional action. One of the widely disputed answers appeals to a type of low-level mental state, motor representations, aiming to bridge the explanatory gap with it, and provide an explanation of what it is to cause an action "in the right way."

However, according to Butterfill & Sinigaglia (2012), theories employing the motor representations are subject to the interface problem: how do the intentions with propositional contents transform into the motor representations with non-propositional, motoric contents? I will review four theories that try to solve the interface problem, thereby the problem of causal deviance in the current literature:

1. The Deferral View form Butterfill & Sinigaglia (2012).
2. The Motor Schema View from Mylopoulos and Pacherie (2017, 2019).
3. Dual Content View from Shepherd (2017a, 2017b).
4. The Same Format View by Ferretti & Caiani (2018, 2019).

In the following, I will begin by further clarifying the interface problem using the DPM model of intentions. I will then introduce the four theories that try to solve this interface problem- outlining the key arguments and objections for each. Next, I will present a case of causal deviance for motor representations that poses a challenge to all four of these theories, under the causalist tradition of trying to establish the right causal chain between intentions and actions, and I will show how they all fail to fully resolve the problem of causal deviance. Finally, I will conclude by hinting at a potential alternative approach, inspired by Wu (2016), that cognitions (i.e., intentions) bias actions rather than causing them directly. This suggests moving beyond the causalist framework to reconceptualize how cognitive states interface with motor representations, while still preserving the key role of motor representations in action production. The overall goal of the paper is to motivate the need for a new theory by revealing how current causalist approaches cannot adequately handle cases of causal deviance for motor representations.

Interface problem, DPM model, and four current theories

I start with introducing the DPM Model, for it helps to clarify the interface problem. The DPM model categorizes the intention of an action into three layers, viz., D(istal)-intentions, P(roximal)-intentions, and M(otor)-intentions, hence the shorthand DPM model (Pacherie, 2006, 2007, 2008; Mylopoulos & Pacherie, 2019).

The D-intention of a moral action is a general form of intentions directed toward the intended goal, e.g., Sam's intention to attend the conference. It will be formed after a specific goal has been determined by the agent and then forms as well as controls the general means of achieving the goal in a rational way. D-intention is also hierarchical structured within the same layer, as it usually, if not always, needs serval more specific D-intentions serving as means to achieve the intended goal. The P-intention implements general goal-directed means inherited from the D-intention in the current situation of action. It forms based on an integration of the general means and the perceptual information of the situation, which allows monitoring and controlling the action a more localized way as it unfolds in the current perceptual situation. Thus, it is a more definitive representation of the action, which includes both the intended goal and more specific means for the situation of action at hand. The M-intention is a motor representation of an action, plus promoting the execution of that action. It stands for the intended goal of an action and the motoric means suitable for direct execution of the action and automatically watches and controls the action on a fine time scale. In this way, the motor representations further decompose the intended goals into motor goals with motoric contents (Grafton & Hamilton, 2007; Jeannerod, 2006; van Elk, van Schie, & Bekkering, 2014).

According to Butterfill & Sinigaglia (2012), while introducing motor representations aids in ensuring ongoing control over the unfolding of actions, a key challenge lies in explaining how intentions with propositional contents can connect and coordinate with motor representations that own motoric, non-propositional formats. Here is how they explain:

> There are cases in which a particular action is guided both by one or more intentions and by one or more motor representations. In at least some such cases, the outcomes specified by the intentions match the outcomes specified by the motor representations. Furthermore, this match is not always accidental. How do non-accidental matches come about? (pp. 131-132)

This leads to what they refer to as the interface problem: how can the cognitive system's representations interface with the motor system's representations when they employ different formats? Here are four theories proposed to solve it:

First, the Deferral View proposed by Butterfill & Sinigaglia (2012): this view draws a parallel between the relationship of propositional to pictorial representations and the relationship of D-intention to motor representations. They argue that it seems that we can use the sentence "follow that route" to refer to a route, where the demonstrative concept "that" (i.e., the propositional representation) refers to the route by means of deferring to the pictorial line of the map (i.e., the pictorial representation). If so, we can also use the D-intention with the propositional contents "do that" to refer to the outcome of an action, where the demonstrative concept "that" (i.e., the intentional representation) refers to the outcome by means of deferring to the motor representations of this action outcome. Then, D-intentions can connect with motor representations via a demonstrative concept within the propositional contents of intentions.

This concept of deferral can sidestep a hypothetical translation between intentions and motor representations, about which Butterfill & Sinigaglia (2012) argue, "nothing at all is known about" and "nor about how it might be achieved, nor even about how it might be investigated" (p.133). However, one critic to Deferral View is that a translation process is still presupposed. The agents must independently know which the *right* motor representations are selected for their D-intentions to defer to. But it would involve a translation process between the D-intentions and the *right* motor representations that are selected, which they have argued is impossible to establish (See detailed citric in Mylopoulos & Pacherie, 2017).

Second, the Motor Schema view proposed by Mylopoulos & Pacherie (2017, 2019): this theory introduce "executable action concepts" that initiates motor schemas serving as an intermediary layer between the P-intentions and motor representations. Note that motor schemas are more stable and abstract motor representations, and fine-grained parameters of action executions that encode the invariant and general features of the actions with respect to their temporal ordering, spatial configurations, relative speeds, and forces (Mylopoulos & Pacherie, 2017, p. 290).

Executable action concepts are formed based on agents' previous action executions in sense that they have the ability to execute the actions. These concepts can function as both contents of P-intentions and directly lead to motor representations of actions via their connections between related motor schemas. Therefore, P-intentions can stipulate the motor schemas in terms of executable action concepts towards motor representations, which cause the motor representations. This, without presupposing any translation process, establishes a content-preserving causal process between the propositional contents of intentions and the motoric contents of motor representations.

However, the critic to the Motor Schema view mirrors that to the Deferral view. (2017a) argues that presupposing "executable action concepts" as concepts with propositional formats that can connect with motor schemas with motoric formats revives the interface problem instead of solving it. It requires agents independently know which motor schemas should be selected for direct employment by motor representations. This seems to require a translation process between intentional representations and motor representations (see also Ferretti & Caiani 2019, p.309).

Third, the Dual Content view proposed by Shepherd (2017a, 2017b): this view posits at least some P-intentions have both propositional and motoric representational contents. As Shepherd (2017a, p.10) writes, "the solution to the interface problem is that intentions lead a double life. Intentions can take propositionally formatted contents that enable their integration with propositional thought. And intentions have motorically formatted contents that communicate in a fairly direct way with the operations of motoric-level action implementation." He argues that this view is supported by empirical evidence regarding intentions and implicit learning in target-approaching tasks. For example, Day et al. (2016) conducted an experiment where participants had to aim at a target on a rotating circle on the screen by moving a cursor. However, the visual feedback provided to the participants was perturbed and non-veridical, meaning it did not accurately stand for the actual position of the hand that controlled the cursor. So, participants had to self-adjust their movements to successfully aim at the target. Before each movement, participants were asked to report the intended location they aimed at. The results showed that participants' actual movements drifted in the direction opposite to the perturbed, non-veridical visual feedback, compensating for the sensory errors. This indicated implicit learning by the motor systems. Furthermore, when participants aimed farther from their often-reported intended locations, the magnitude of implicit learning decreased. This suggested that intentions set the locations where implicit sensorimotor adaptations occur. Collectively, these findings prove a close interaction between D-intentions and motor representations, lending support to the Dual Content view.

A major concern with this view is that while the empirical evidence directly supports a close relationship between intentions and motor representations, it only indirectly supports the Dual Content view. The Dual Content view claims more than just a close relationship, proposing that D-intentions have a combined propositional-motoric format, rather than merely suggesting a tight coupling between the two components. Perhaps it would be more proper for Shepherd to make a more moderate claim about the connection between intentions and motor representations. As he argues (2017a), cognitions guide motor representations rather than keeping a strict causal connection with them, which better aligns with the empirical evidence he presents. This moderate stance, proposing that cognitions guide motor representations without making stronger claims about their combined format, may be a more fitting interpretation of the empirical findings. Perhaps it would be more proper for Shepherd to adopt a more moderate stance on the connection between intentions and motor representations, akin to his argument (2017b) that cognitions guide motor representations rather than staying in a strict causal connection with them. This moderate claim of cognitions guiding motor representations, without making stronger claims about their combined format, better aligns with the empirical evidence he presents and may be a more fitting interpretation of the results.

Fourth, the Same Format view proposed by Ferretti & Caiani (2018). They argue that D-intentions and motor representations share the same non-propositional, motoric format, allowing them to interact directly. Three groups of behavioural and neurobiological evidence seem to support this claim. First, some propositional representations and action executions influence each other bi-directionally, with the former modulating the latter, and impairments in the latter affecting the formation of the former. For example, Gentilucci and Gangitano's study (1998) showed that when participants reached out and grasped rods labelled "long" or "short," the words affected their arm movement parameters. Second, they rely on the grounded cognition hypothesis, which states that propositional representations are deeply grounded in sensorimotor systems. For example, Hauk et al.'s study (2004) found that reading verbs involving hand, foot, and mouth movements elicited activation in corresponding motor regions of the brain. Third, propositional representations of action are argued to be part of the motor systems' activity. For example, Buccino et al.'s study (2005) used TMS and found that listening to action-related sentences modulates motor system activity, particularly in muscle groups associated with the action that is mentioned in the sentence.

However, it seems that for agents to independently know which propositional representations the right ones are to share the motoric format with motor representations, a translation process is needed. As Christensen argues "they tacitly abandon the assumption of no translation, and it is therefore not clear why intentions and motor representations need to be in the same format." (2020, p.548), One plausible account is that the translation process might already be implicit in the Same Format view when they argue that some propositional contents can be represented in motor systems. If so, it simply transforms the interface problem into a new form: how is it possible for intentions to interact with motor representations within the motor systems?

All four theories discussed above hold that intentional actions can be constitutively caused by propositional attitudes "in the right way," by virtue of motor representations serving as mediators in different but similar senses. They are considered causalist views of intentional action. While they provide innovative and promising insights into the interface problem, each faces internal challenges that either render the theory itself flawed or cause it to fall back into the same problem about a mysterious translation process. In the next section, I will argue that even if we grant that these theories succeed in offering an account with the help of motor representations, they still fall short of solving the problem of causal deviance.

Cases of causal deviance for motor representations

I will now argue that in addition to the internal objections each of the above four views faces, there is a further type of objection that none of them can adequately respond to, rendering it even more devastating. As we have seen, all four views attempt to set up a causal process between intentions and the motor representations of actions, with the goal of eliminating the accidental connection between intentions and their caused events, thereby addressing the problem of casual deviance. However, even if such a connection exists, it cannot guarantee the non-accidental nature of the connection because it is easy to introduce accidental components into the connection. The following cases will demonstrate this point.

As long as motor representations are well preserved in a deviant causal chain, it is always possible to devise situations where:

a. The agent intended to execute action A

b. The outcome of their intending to execute A corresponds to the agent’s intentional representation of what will happen.
c. The outcome of their intending to execute A corresponds to the agent's motor representations of what will happen.
d. The outcome of their intending to execute A, while caused by their intending to execute A, should be considered accidental rather than intentional.

To illustrate this, let us consider several cases involving different actions and agents.

Case 1: Proposing

John intended to get down on one knee to propose marriage. However, as he arrived in front of his beloved, a sudden wave of cramps overtook him, brought on by the overwhelming stress he had endured. These cramps directly caused John to lower himself to his knee. Although the action appeared smooth, and John was unaware of the cramps. His motor representation of getting down on his knee was so normal that it did not detect any disruption in the action. To John’s intention and motor representations, his kneeling remained part of his intended causal process.

Here, the cramps, not John's intention, were the direct cause of his lowering to one knee. This case proves how the current four theories fail to account for why having the correct intention combined with the appropriate motor representations does not necessarily guarantee that the resulting outcome was caused by an intentional action. Let us examine how each of the four current theories would respond in Case 1 and see why they do not provide a satisfactory explanation.

First, the Deferral view posits that intention defers to motor representations. In Case 1, let us assume John also has the intention “do that,” as presupposed by the Deferral View, where “that” refers to the action of getting down on his knee. According to this view, this intention should defer to the motor representations of unfolding of that action. However, John already had the appropriate motor representations for his action, which the intention can successfully defer to.

Secondly, the Motor Schema View proposes that “executable action concepts” initiate motor schemas, serving as intermediaries between intentions and motor representations. Presumably, John could have these unique concepts of getting down on his knee, capable of starting the motor schemas constituting the action in question. His motor systems might store the motor schemas for that action from executing similar actions, such as "getting down to pick something up." If so, it seems to meet the criteria of the Motor Schema View's account of intentional actions.

Thirdly, the Dual Content view posits that some intentions have both propositional and motoric representational contents. Suppose John's intention of getting down on his knee has a propositional representational content as well as motoric representational content involving motor representations for the movements needed to kneel. However, there seems to be nothing unusual about the motor representations of John's action in Case 1 – it is no different than the motor representations involved when he intended to do so.

Fourthly, the Same Format View proposes that D-intentions and motor representations share the same non-propositional, motoric format. John's motor representations are just as right as they should be. Thus, the criteria of the Same Format view's account of intentional actions can be met. However, John ended up kneeling because of the cramp, not because of his intention. This case proves how the current four theories fail to account for why having the correct intention combined with appropriate motor representations does not necessarily guarantee that the resulting outcome was intentional.

One might question whether it is truly plausible for John to have normal motor representations in this case. After all, cramps typically disrupt normal motor control, leading to uncoordinated or spasmodic movements. However, research in motor control and neurology has shown that in certain high-stress situations, individuals can experience involuntary muscle contractions that mimic intentional actions, particularly when those actions are highly rehearsed or routine. For example, in cases of dystonia—a neurological disorder characterized by involuntary muscle contractions—patients can sometimes perform complex motor tasks with apparent normalcy, even though the underlying cause is an involuntary movement. This suggests that it is possible for John's motor representations to remain intact, despite the cramps being the primary cause of his kneeling.

Critics might argue that a causal history exists between his cramp and intention. For example, his intention to propose causes stress, which causes cramps, and the cramps cause appropriate motor representations controlling the successful unfolding of his actions. In this sense, the cramp is not merely accidental for John but necessary to execute such action appropriately. If so, a specific cramp is always involved in the causal process of his action, in the sense that every time he intends to propose, such a cramp will occur to ensure he successfully gets on knee. Imagine an old painter with Parkinson's disease who cannot control his hands, they just keep shaking. But every time he paints, his hand with the brush does not shake because his attention is fully centred on the painting, causing an effect on his neural systems that override what is causing his hand to shake. The effect, then, is a necessary factor in the causal process of executing the action (i.e., using the brush without shaking it), without which the action cannot unfold properly.

However, this response might overlook the importance of distinguishing between what is considered an intentional action and what is merely a successful outcome. Just because the outcome appears intentional does not mean that the action itself was. The crucial point here is that John's motor representations, while appearing normal, do not alter the fact that the cramps were the actual cause of the action. This case illustrates how normal motor representations can be involved in actions that are unintentional due to deviant causal chains.

But I do not think it is the right response to explain the cramp as a factor in the causal process. It cannot account for the casual deviance when minor changes are made to Case 1. Consider a case as follows:

Case 2: Presentation

Jane intended to raise her hand to ask a question during a presentation. However, as she prepared to do so, she experienced an involuntary muscle spasm caused by a newly developing neurological condition. This spasm caused her arm to rise in a manner indistinguishable from her usual hand-raising motion. Jane's motor representations of the action were normal, and she was unaware of the spasm's influence.

The neurological condition causing the spasm is independent of Jane's intentions, blocking the reply that worked for Case 1. This demonstrates that even when motor representations are preserved, the action can still be unintentional due to a deviant causal chain. Unlike Case 1, where one might argue that the cramp was indirectly linked to the intention through stress, Case 2 shows that the spasm is entirely unrelated to Jane's intention, thus blocking the response that could be given for John's case. Therefore, Jane's action of raising her hand was not intentional, despite being caused by her intention and corresponding to her motor representations.

Here, too, one might question the assumption of normal motor representations. While it might seem improbable that a neurological condition could produce such a precise, intentional-like movement, there is evidence in neurology that certain involuntary movements can be highly specific and controlled. For instance, in cases of focal dystonia, individuals can experience involuntary muscle contractions that replicate intentional movements, especially in contexts where the movement is well-practiced or habitual. This evidence supports the possibility that Jane's motor representations could remain normal despite the underlying involuntary spasm, reinforcing the argument that preserved motor representations do not necessarily equate to intentional actions.

Consider Case 3 to further illustrate this point.

Case 3: Golf Swing

Jun intended to swing his golf club to hit the ball. As he prepared to swing, his anxiety about his performance triggered a psychosomatic response, causing his muscles to tense and relax in a pattern that perfectly mimicked his usual golf swing. Jun's motor representations of the swing were normal, and he was unaware of the psychosomatic influence on his action.

Jun's psychosomatic response, while triggered by Jun's mental state, created a deviant causal chain between his intention and the action. This case further illustrates how preserved motor representations do not guarantee intentional action. Just as in Jane's case, the psychosomatic response is not directly related to the intention to swing, thereby reinforcing the point that even with intact motor representations, the action can still be unintentional due to causal deviance. Therefore, Jun's golf swing in this case, while caused by his intention and matching his motor representations, was not intentional.

Before presenting a general diagnose of these cases, I want to address the question of why we could expect normal motor representations in such cases. Motor representations are proved to be generated and executed unconsciously, without direct conscious control. This means that even in cases of causal deviance, the underlying motor systems may continue to function normally. And the human body has many feedback systems that continuously adjust and refine movements. These systems may compensate for unusual inputs, resulting in apparently normal motor representations. Also, well-practiced actions often become automatic, with motor representations stored in procedural memory. These ingrained patterns may persist even when the causal chain is disrupted.

My diagnosis is that the motoric format of the motor representations contributes to its bottom-up nature. However, the top-down causal chains from intentions cannot necessarily guarantee the right type of motor representations is non-accidentally caused by them. Moreover, when an action execution unfolds smoothly and successfully (e.g., John is getting down on his knee appropriately), motor representations just automatically execute without "questioning" whether they were caused by intention in the right or deviant manner. In this sense, motor representations of actions are outcome-oriented -- they are not responsible for assuring the right causal chain as long as the outcome is right, namely, the action is executed properly.

If this diagnosis is correct, then all four of the theories discussed earlier still do not provide an adequate answer to the problem of causal deviance. Why do they fail? One plausible explanation is that they have not solved the interface problem. I do not deny the plausibility. As discussed in the last section, many attempts have been made in this direction. Another equally plausible explanation is that the motor representations cannot rule out the accidental factor in the causal chain. The challenge remains to specify the right causal chain that must exist between intentions and the outcomes of the action in order for the latter to qualify as intentional actions. Adding motor representations to the causal chains of intentions does not help causalists solve the problem of causal deviance.

Conclusion

In conclusion, I have critically examined four prominent theories that propose motor representations as a solution to the problem of causal deviance in theories of intentional action. While providing insightful perspectives, each theory faces internal challenges, and more importantly, none can adequately account for cases where the motor representations link intentions to action outcomes, yet the outcome is still accidental rather than intentional.

A key diagnosis is that motor representations, by virtue of their motoric, bottom-up nature, are primarily outcome-oriented rather than ensuring the right causal chain from intentions. As long as the outcome matches the motor representations, they will execute the associated action regardless of whether the causal chain was deviant or not. The theories explored all operate within a causalist framework, attempting to establish the right kind of causal chain between intentions and motor representations of action. However, the cases of causal deviance for motor representations suggest that even with motor representations in the picture, unintentional "mismatches" can still arise, indicating the persistence of the problem of causal deviance.

Moving forward, these challenges motivate exploring alternatives that go beyond the current causalist frameworks. One possibility, inspired by Wu (2016), is to view cognitions like intentions as biasing rather than strictly causing motor representations and actions. Intentions modulate attention to bias the coupling of goal-directed perceptual information and goal-directed execution actions. The attention mechanism selects and highlights appropriate targets for action to execute on. The perceptual information prioritized by attention is then used in forming the relevant motor representations that control actions. This biasing view proposes that the rich interactions among intentions, attention, and the perception-action coupling processes explain how specific intentions arise and propagate to shape the motor representations controlling and monitoring the unfolding of actions. This could account for the tight coordination between cognition and action in empirical studies about attention, while avoiding the causalist commitment of intentions directly causing actions in a predetermined manner.

Rethinking how cognitive states interface with motoric representations, outside a causalist model, may offer a more promising path toward avoiding instead of solving the problem of causal deviance. The overall moral is that the intentional actions is not a product of being caused by prior propositional attitudes, and motor representations alone cannot be the full answer – a reconceptualization of the cognition-motor interface itself is needed.

References

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–363. https://doi.org/10.1016/j.cogbrainres.2005.02.020

Bratman, M. (1984). Two Faces of Intention. *The Philosophical Review*, *93*(3), 375. https://doi.org/10.2307/2184542

Bratman, M. (1987). *Intentions, plans, and practical reason*. Harvard University Press.

Butterfill, S. A., & Sinigaglia, C. (2012). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*, *88*(1), 119–145. https://doi.org/10.1111/j.1933-1592.2012.00604.x

Christensen, W. (2020). The Skill of Translating Thought into Action: Framing The Problem. *Review of Philosophy and Psychology*, *12*(3), 547–573. https://doi.org/10.1007/s13164-020-00517-2

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*(23), 685–700. https://doi.org/10.2307/2023177

Davidson, D. (2002a). Agency, In D. Davidson, *Essays on Actions and Events* (pp. 43-61). Oxford University Press. (Reprinted from *Agent, Action, and Reason*, pp. 1-37, by R. Binkley, R. Bronaugh, and A. Marras, Eds., 1971, University of Toronto Press)

Davidson, D. (2002b). Intending. In D. Davidson, *Essays on Actions and Events* (pp. 83-102). Oxford University Press. (Reprinted from *Philosophy of History and Action*, pp. 41-60, by Y. Yovel, D. Reidel, Eds., 1978, The Magnes Press)

Day, K. A., Roemmich, R. T., Taylor, J. A., & Bastian, A. J. (2016). Visuomotor Learning Generalizes Around the Intended Movement. *ENeuro*, *3*(2), ENEURO.0005-16.2016. https://doi.org/10.1523/ENEURO.0005-16.2016

Ferretti, G., & Caiani, S. Z. (2018). Solving the Interface Problem Without Translation: The Same Format Thesis. *Pacific Philosophical Quarterly*, *100*(1), 301–333. https://doi.org/10.1111/papq.12243

Ferretti, G., & Caiani, S. Z. (2019). A model for the interlock between propositional and motor formats. In M. Fontaine, C. Barés-Gómez, F. Salguero-Lamillar, L. Magnani, & Á. Nepomuceno-Fernández (Eds.), *Model-Based Reasoning in Science and Technology: Inferential Models for Logic, Language, Cognition and Computation*. Springer Verlag.

Gentilucci, M., & Gangitano, M. (1998). Influence of automatic word reading on motor control. *European Journal of Neuroscience*, *10*(2), 752–756. https://doi.org/10.1046/j.1460-9568.1998.00060

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/s0896-6273(03)00838-9

Mylopoulos, M., & Pacherie, E. (2017). Intentions and motor representations: The interface challenge. *Review of Philosophy and Psychology, 8*(2), 317–336. https://doi.org/10.1007/s13164-016-0311-6

Mylopoulos, M., & Pacherie, E. (2019). Intentions: The dynamic hierarchical model revisited. *WIREs Cognitive Science, 10*(2), e1481. https://doi.org/10.1002/wcs.1481

Pacherie, E. (2006). Towards a dynamic theory of intentions. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation of the nature of volition* (pp. 145–167). Cambridge, MA: MIT Press.

Pacherie, E. (2007). The sense of control and the sense of agency. *Psyche, 13*(1), 1–30. https://hal.science/ijn_00352565v1

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*(1), 179–217. https://doi.org/10.1016/j.cognition.2007.09.003

Paul, S. (2009). How we know what we're doing. *Philosopher's Imprint, 9*(11), 1–24. http://hdl.handle.net/2027/spo.3521354.0009.011

Shepherd, J. (2017a). Skilled action and the double life of intention. *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12433

Shepherd, J. (2017b). The experience of acting and the structure of consciousness. *The Journal of Philosophy, 114*(8), 422–448. https://doi.org/10.5840/jphil2017114829

Wu, W. (2016). Experts and deviants: The story of agentive control. *Philosophy and Phenomenological Research, 93*(1), 101–126. https://doi.org/10.1111/phpr.12170

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 94-119

Artículo/Article

Semantics of the liar paradox

Semántica de la paradoja del mentiroso

C. P. Hertogh

Free University Brussels –(VUB-BE/EU)
noromyxo2005@gmail.com
https://orcid.org/0000-0002-6778-4837

DOI https://doi.org/10.48204/2805-1815.8475

INFORMACIÓN DEL ARTÍCULO

Recibido el:12/1/2025

Aceptado el: 22/8/2025

Keywords:

Excluded middle, non-contradiction, paraconsistent logic, semantics, truth

Palabras clave:

Medio excluido, no contradicción, lógica paraconsistente, semántica, verdad

ABSTRACT/RESUMEN

Abstract:

Although Liar-type conundrums –traditionally considered sophistry– do not match all characteristics of thought experiments (TE), particularly not the pragmaticist condition that thought experiments are designed to resolve predefined problems, we apply TE analyses and interpretations.

The *Liar* (or, let's say, *Liar*-type statements involving truth (predicates), self-reference as in the fields of conceptual analysis, semantics and set theory) rose to paradigmatic, revolutionary prominence by Tarski's Gödelian logicistic deliberations in the beginning of last century now considered the orthodox semantic account.

We survey semiotic and pragmatic accounts from the second half of last century and (non)classical (meta)logical accounts that may gain traction in 21st century.

Our resolution is manifold, both semantic and pragmatic. We show that modern logic, from the very beginning in Wittgenstein, has had more than two truth-values, next to T(rue) and F(alse), 'nonsense', 'meaningless', 'senseless', etc.

We show that the *Liar* may bring up for discussion logical principles like law of excluded middle (LEM) and noncontradiction (LNC) and refute strong versions of these logical laws. We propose a pragmatic Gricean account of the *Liar*, analyzing Epimenides's and Eubulides' versions of the Liar as falsification of generic tacit conversational principle –Grice's maxim of quality and maxim of manner– that people usually speak truthful, speak perspicuously.

By application of TE Matrix Epimenides-style and Eubulides-style paradoxes can be extended to valid and sound modus tollens-style arguments with logical force of enthymeme.

Resumen:

Aunque los enigmas como la paradoja del mentiroso —tradicionalmente considerados sofistería— no se ajustan a todas las características de los experimentos mentales (TE), en particular a la condición pragmaticista de que los experimentos mentales están diseñados para resolver problemas predefinidos, aplicamos análisis e interpretaciones de TE.

La paradoja del mentiroso (o, digamos, las afirmaciones de tipo mentiroso que involucran, verdad (predicados), autorreferencia, como en los campos del análisis conceptual, la semántica y la teoría de conjuntos) alcanzó una prominencia paradigmática y revolucionaria gracias a las deliberaciones logicistas gödelianas de Tarski a principios del siglo pasado, ahora consideradas la explicación semántica ortodoxa.

Examinamos las explicaciones semióticas y pragmáticas de la segunda mitad del siglo pasado y las explicaciones (meta)lógicas (no)clásicas que podrían cobrar fuerza en el siglo XXI.

Nuestra resolución es múltiple, tanto semántica como pragmática. Demostramos que la lógica moderna, desde sus inicios, Wittgenstein ha tenido más de dos valores de verdad, además de Verdadero y Falso, 'sinsentido', etc.

Mostramos que la paradoja del mentiroso puede llevar a replantear principios lógicos como la ley del tercero excluido (LEM) y la no contradicción (LNC), y refutamos versiones fuertes de estas leyes lógicas. Proponemos una explicación pragmática griceana de la paradoja del mentiroso, analizando las versiones de Epiménides y Eubulides como una falsificación del principio conversacional tácito genérico —la máxima de calidad y la máxima de modo de Grice— según el cual las personas suelen hablar con la verdad y con perspicacia.

Mediante la aplicación de la Matriz TE, las paradojas de Epiménides y Eubulides pueden extenderse a argumentos de estilo modus tollens válidos y sólidos con la fuerza lógica del entimema.

Identifications

The *Liar* appears in different versions throughout history of logic and philosophy (for recent discussions, see Martin, 1974, 1984; Beall et al. 2016/2023). We will particularly focus on semantic (instead of set theoretical) versions, such as Epimenides (7th-6th cent. BCE) and Eubulides (4th cent. BCE) versions.

Epimenides-type *Liar* paradoxes,

The Cretan says, 'All Cretans are liars'. [a]

The paradox is also in the Bible, New Testament, Titus I, 12 (Kripke, 1975, p. 690), *NABR* (*New American Bible (Revised Edition)*):

"One of them, a prophet of their own, once said, 'Cretans have always been liars, vicious beasts, and lazy gluttons'". [b]

From point of view of present-day Critical Thinking courses, it appears to be a strongly emotional utterance of a negative prejudice against Cretan people, which bias should be removed, after which the proposition may have lost most of its cognitive content, too. It is one more reason not to allow for full range application of universal quantifier, since it seems to be an emotional exaggeration and for that reason already only partly true or not true at all.

Nevertheless, we will discuss and analyse this version in a foremost logical way as the logical paradox it has been turned into by history of logic and philosophy.

Eubulides-type *Liar* paradoxes,

A man says: "What I am saying now is a lie". [c]

This statement is false. [d]

The *Liar* or *pseudomenos* (*ψευδόμενος*) paradox is well-known in logic as a false, sophistical species of syllogism. Eubulides is also mentioned as conceiver of some more sophistic paradoxes, such as *The Heap* and *The Bald Man*. Traditionally, and postmodernistically again, these paradoxes are discussed in logic and linguistics as about vague or ambiguous predicates.

For the mathematical modern set theoretic versions (as *Russell's Paradox*, semantic version called *The Barber*) that aim at developing axiomatized logical systems, a more formal and abstract account may be required, which may be applied in development of computer languages, computer science, AI (Artificial Intelligence), IT (Information Technology).

We will survey some semantic historical accounts and then propose a semantic, pragmatic account touching here and there on insights from formal logical analyses of natural language, applying Chomsky linguistics (Chomsky, 1957/2002) and Kripke semantics (Kripke, 1975, 1980, 1982, 2019),

Fourfold Definition

We propose a four-fold definition of TE

1. Ontologically speaking, TE are mental phenomena.
2. After linguistic and hermeneutic turn, TE are TE texts
3. Logically speaking, TE are incomplete arguments (enthymemes).
4. On pragmaticist view, TE are mental devices designed to resolve predefined problems (analogous to experiments in science). (See also, Hertogh, 2015, 2018, 2023a on Galileo, 1638/1914; Hertogh, 2024 on Einstein & Infeld, 1938).

Traditionally considered sophistry the *Liar* does not live up to the fourth condition, since the *Liar*-type statements were not designed to resolve predefined problems, but rather to show limits of logic.

Also, they miss TE-indicators as 'Consider …', 'Imagine …', etc. Nevertheless, we may try and apply TE analyses and interpretations.

Crucial Thought Experiments

The *Liar* rose to logical prominence with Tarski's Gödelian (1933/1935) account. Tarski considers the *Liar* crucial experiments, which should be crucial *thought* experiments, since it is not about empirical sciences but mathematics and logic, which experiments don't need confirmation by execution in external reality since the basic entities of mathematics and logic only exist in the mind as ideal entities or nowadays in virtual space (as line, circle, numbers), which can only be physically approximated on Planet Earth.

> Personally, as a logician, I could not reconcile myself with antinomies as a permanent element of our system of knowledge. However, I am not the least inclined to treat antinomies lightly. The appearance of an antinomy is for me a symptom of disease. Starting with premises that seem intuitively obvious, using forms of reasoning that seem intuitively certain, an antinomy leads us to nonsense, a contradiction. Whenever this happens, we have to submit our ways of thinking to a thorough revision, to reject some premises in which we believed or to improve some forms of argument which we used. We do this with the hope not only that the

ISSN – L 2805 – 1815

> old antinomy will be disposed of but also that no new one will appear. To this end we test our reformed system of thinking by all available means, and, first of all, we try to reconstruct the old antinomy in the new setting; this testing is a very important activity in the realm of speculative thought, akin to carrying out crucial experiments in empirical science. (Tarski,1969, pp. 4-5)

Possibly the *Liar* could be considered kind of consistency TE test.

Mathematical Thought Experiments

Updated modern versions of the *Liar* may be considered paradigmatic or revolutionary thought experiments, as arising in 'foundational crises' contributing to theory change, a 'new conceptual framework'.

In Horowitz's & Massey's (1991) pivotal collection of essays on thought experiments in philosophy and science, Dionysios Anapolitanos contributes an essay on thought experimentation and conceivability conditions in mathematics, in which set theoretical paradoxes are discussed both as incentives to TE and TE themselves,

> The third group includes thought experiments in mathematics performed fervently during and immediately after a foundational crisis. The overall activity during such periods is mainly centered around the construction of a new conceptual framework wherein the source of crisis in the old framework is hoped to be tamed… The best and the most well-known example of such creative activity spurred by a foundational crisis is the one occurred at the beginning of the 20th century during and after the appearance of the set-theoretic paradoxes. Various proposals concerning the modification of the naive Cantorian concept of set were put forwards with some notable among them those of the Russellian theory of types, of the Zermelo-Fraenkel theory of sets and of the Gödel-Bernays set theory… All these proposals not only *started* as thought experiments, but they *were* thought experiments in the sense of open-ended explanatory attempts in a playful and in a state of crisis conceptual framework, which, so to speak, was set free in motion by the emergence of the paradoxes. (Anapolitanos, 1991, p. 93)

An example of a set theoretic paradox in mathematical logic is so-called *Russell's Paradox*, which comes together with popular version of *Barber Paradox*, which is akin to Epimenides's version of the *Liar* asking 'whether a class is a member of itself or not', which may yield a contradiction "... the question whether the barber shaves himself or not. You can define the barber as 'one who shaves all those, and those only, who do not shave themselves.' The question is, does the barber shave himself?" (Russell, 1919 , pp. 354-355)[1], which argument Russell concludes to 'nonsense'

> ... But in our previous form I think it is clear that you can only get around it by observing that the whole question whether a class is or is not a member of itself is nonsense, i. e., that no class either is or is not a member of itself, and that it is not even true to say that, because the whole form of words is just a noise without meaning. (Russell,1919, p. 355)

Wittgenstein proposed a resolution in Tractatus 3.333, referring to the functions variant rather than the classes variant of the paradox, arguing a function cannot be an argument of itself nor contain itself, possibly demonstrating it as by adding existential quantification (∃ϕ) over argument (u): "This is at once clear, if instead of 'F(F(u))' we write '(∃ϕ): F(ϕu) . ϕu = Fu'. Herewith Russell's paradox vanishes", (Wittgenstein, 2022, #3.333) which formulas, however, may appear equivocal on interpretation, even involving discussions on a possible misprint in some editions of Tractatus (see Black, 1964; Jolley, 2004; Sutrop, 2009). We will come back to it in section about senselessness (Beyond Two Truth Values).

Paradigmatic Thought Experiments

The terms *paradigmatic* or *revolutionary* TE are derived from Kuhn's *The Structure of Scientific Revolutions.* Focusing on TE in physics, Kuhn holds that paradigmatic TE contribute to theory change in times of foundational crisis:

> It is no accident that the emergence of Newtonian physics in the seventeenth century and of relativity and quantum mechanics in the twentieth century should have been both preceded by and accompanied by fundamental philosophical analyses of the contemporary research tradition. Nor is it an accident that in both these periods the so-called thought experiment should have played so critical a role in the progress of research. As I have shown elsewhere, the analytical thought experimentation that bulks so large in the writings of Galileo, Einstein, Bohr, and others is perfectly calculated to expose the old paradigm to

existing knowledge in the ways that isolate the root of the crisis with a clarity unattainable in the laboratory. (Kuhn, 1970, p, 88)

Paradoxes

We may define paradoxes like this: Paradoxes are defined in this paper as metalogical instruments that question their very presuppositions or the presuppositions of the (logical) system they are part of.

Like TE they may be resolved by Extended Argument analyses, i.e., by explicitation of rules, principles, presuppositions that validate them as (logical) arguments, within a different or more encompassing (logical) system.

They are enthymemes in senso stricto when it is about finding a major under which they may subsume as minor. The missing premises, presupposition, etc., that make them come true, may be found in immediate context or theory (immanent or internal analyses), or outside immediate context or theory (transcendental, possibly transcendent or external analyses). This division may partly coincide with division between *weak* and *strong paradoxes* and between *seeming* and *apparent paradoxes.*[2]

According to Wójtowicz (2021) paradoxes hold an important place in philosophy—they force us to verify our beliefs and inspire us to search for new solutions. Still, there is a difference with TE since the paradoxes were not designed to resolve predefined problems and the conceivers may have been perplexed by them as well, so, they only conceived of new problems, new phrasings of problems instead of conceiving of resolutions for existing problems and they increased the number of problems instead of decreasing them.

Nevertheless, some philosophers define TE with reference to paradoxes, and we list the paradox view among picks of views in 2015 dissertation. Sorensen (1992) advances some theories, types and functions of TE, and his main theory seems to center around paradoxes "... A paradox is a small set of individually plausible yet jointly inconsistent propositions … every thought experiment is reducible to such a set. … They are the molds in which raw TE can be poured. They then enter the logician's mill...". (Sorensen, 1992, p. 122)

However, Sorensen does not discuss the *Liar* in his main (1992) work on TE, it does not appear in the subject index, nor is Tarski discussed or does Tarski appear in the name index.

Sophistry, Huàtóus and Kōans

In earlier texts we have stressed possible ill intent of paradoxes, particular sophisms as used by sophists in ancient Greece (see Kirk et al., 1983) a as they were made and intended to deceive (e.g. *Achilles and the Tortoise*), especially in contrast to *Huàtóus* and *Kōans* that have a predefined spiritual sense (that may have beneficial effects on health as well) and for that reason match out definition of TE better. (See Hertogh, 2018, on *Vipassanā Meditation* and Hertogh, 2021b, on *Huàtóus* and *Kōans*.)

Analyses

Orthodox Semantic Accounts

The *Liar* (or, let's say, *Liar*-type statements involving truth (predicates), self-reference as in the fields of conceptual analysis, semantics and set theory) rose to paradigmatic, revolutionary prominence by Tarski's Gödelian logicistic deliberations in the beginning of last century (Tarski 1933/1935,1983a, 1983b; Gödel, 1931, 2000), now considered the orthodox semantic account. This account remains within the correspondence theory (e.g., Aristotle, Augustine, Aquinas, Wittgenstein *Tractatus*). The *Liar*-type problem in these accounts does not consist in resolving the paradox, but in revising the logical system in such a way that the paradox will not occur anymore in the revision. It yielded Tarski's Convention T— 'p' [is true] iff p (e.g., 'ravens are black' [is true] if and only if ravens are black) (Tarski, 1933/1935, 1944, 1969, 1983b).

Text Analyses and Speech Act Accounts

Rise of semiotics and pragmatics in the second half of last century added textual analysis (Barthes, 1973, 2014, e.g. 'I am dead', derived from his analyses of an Edgar Allen Poe's story) and speech act accounts (Kearns, 2007; Epstein, 2015, Hoinarescu, 2018, lying as performative or counterformative speech act), revealing naturalist and ideal language biases.

Nonclassical Logical Accounts

Now nonclassical logical accounts may gain traction, revealing limitations and biases of standard logic. Matching our extended TE analyses, we mention some recent paracomplete and paraconsistent approaches to the *Liar*, which assume that the Liar may bring up for discussion logical principles like law of excluded middle (LEM $P \vee \neg P$ —Kripke, 1975, Beall at al. 2016/2023) and noncontradiction (LNC $\neg(P \wedge \neg P)$ —Priest,1984; Karačić 2019), although we have to add that Kripke 1975 (e.g., $\exists x\ (Px \rightarrow Qx) \wedge Qx$ is $\neg T$) may want to be considered classical logic, possibly grounding logic in natural language, intuitions and kind of empirical observation sentences (Carnap, 1928, 1967), facts, state of affairs (Wittgenstein 1922, 1961). etc.

Beall et al. (2016/2023) argue that the Liar has formed 'the core of arguments against classical logic'--arguments for paracomplete logics (e.g. Kripke, 1975; Field, 2008) and paraconsistent logics (Asenjo, 1966; Priest,1984, 2006).

Beall et al. (2016/2023) discusses Kripke's *Liar* theory as 'most influential' example of a paracomplete approach to the *Liar*, where LEM fails 'ín some sense'. *Liar* sentences are neither true (T) nor false (F), but Kripke himself does not use an epithet like 'paracomplete' for his theory and writes about 'truth- value gaps' instead.

Priest (1979, 1984) proposes to accept the paradoxes as from a paraconsistent view on logic,

> The purpose of the present paper is to suggest a new way of handling the logical paradoxes. Instead of trying to dissolve them, or explain what has gone wrong, we should accept them and learn to come to live with them. …For obvious reasons this will require the abandonment, or at least modification, of 'classical' logic... (Priest, 1979, p. 220)

Priest suggests accepting 'some sentences are true (and true only), some false (and false only), and some both true and false!' (Priest, 1979, p. 220).

The first version of the logic of paradox may already have been stablished in 1966 by Argentinian philosopher Florencio González Asenjo. Paraconsistent logics (term coined by Peruvian philosopher Francico Miró Quezada Cantuarias in 1976), and logics developed by Asenjo and Brazilian logician Newton Da Costa (Da Costa, 1974), don't hold on to LNC in classical sense of 'ex contradictione sequitur quodlibet' ('from a contradiction, anything follows'), principle of explosion, but do allow for inconsistencies if not explosive and leading to triviality.

Saul Kripke

According to Kripke:

> The versions of the Liar paradox which use empirical predicates already point up one major aspect of the problem: *many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features...* (1975, p. 691)

Kripke does not mention many examples of 'extremely unfavorable' 'empirical facts' 'to exhibit paradoxical features,' (apart from 'Suppose, however, that Nixon's assertions about Watergate are evenly balanced between the true and the false, except for one problematic case', with regard to Nixon's Watergate scandal in 1972- 1974), but as Kripke advances a mathematical example of possible world (Kripke, 1980, 16--'The thirty-six possible states of the [two] dice are literally thirty-six 'possible worlds,...'--see Hertogh 2021a), one might think of Quine's near-mathematical example of Frederic, reaching age of 21 after passing only five birthdays—since Frederic was born on February 29 (Quine, 1976, p. 1).

'Seeking alternatives to the orthodox approach' (Kripke, 1975, pp. 698-700) since paradoxical sentences are 'ungrounded', they have 'truth-value gaps' and only 'partially defined predicates', choosing for Kleene's 'strong three-valued logic' (1952): Let us suppose that $\neg P$ is true (false) if P is false (true), and undefined if P is undefined.

However, in a note, Kripke (1975, pp. 700-701, note 18) adds the term "three-valued logic'" could be misleading and "our considerations can be formalized in a classical metalanguage",

> ... I have been amazed to hear my use of the Kleene valuation compared occasionally to the proposals of those who favor abandoning standard logic ' for quantum mechanics,' or positing extra truth values beyond truth and falsity, etc. Such a reaction surprised me as much as it would presumably surprise Kleene, who intended (as I do here) to write a work of standard mathematical results, provable in conventional mathematics. ' Undefined' is not an *extra* truth-value …. Nor should it be said that ' classical logic' does not generally hold .... The term ' three-valued logic,' occasionally used here, should not mislead. All our

considerations can be formalized in a classical metalanguage.

'Alternate Intuition'

According to Kripke (1975), paradoxes may be interpreted within a classical metalogical framework on at least an 'alternate intuition', which 'arises only after we have reflected on the process embodying the first intuition',

> The approach adopted here has presupposed the following version of Tarski's 'Convention T', adapted to the three- valued approach: If '**k**' abbreviates a name of the sentence *A, T* (**k**) is false, respectively iff *A* is true, or false. This captures the intuition that *T*(**k**) is to have the same truth conditions as *A* itself; it follows that *T*(**k**) suffers a truth- value gap if *A* does. An alternate intuition would assert that, if *A* is either false or undefined, then *A* is not *true* and *T*(**k**) should be *false,* and its negation *true.* <u>On this view, T(x) will be a totally defined predicate and there are no truth-value gaps.</u> Presumably Tarski's Convention T must be restricted in some way. (Kripke,1975, pp. 714-715 underline added)[3]

In note 35 he explains: "… I think the primacy of the first intuition can be defended philosophically, and for this reason I have emphasized the approach based on this intuition. The alternate intuition arises only after we have reflected on the process embodying the first intuition". (Kripke, 1975, p. 715, note 35)

Also, Ripley (2013) wants to hold on to classical logic by extension of classical logic with 'a fully transparent truth predicate' and 'fully tolerant vague predicates'.[4]

Senselessness Etc. (Beyond Two Truth-values)

"I am somewhat uncertain whether there is a definite factual question as to whether natural language handles truth- value gaps- at least those arising in connection with the semantic paradoxes-..." (Kripke, 1975, p. 712) There are many more problems as in formalization of natural language that have triggered *extensions* of the logical systems *but not revisions,* such as:

A. Meaningless, senseless, nonsensical sentences, e.g., example of grammatical but 'nonsensical' sentence. "Colorless green ideas sleep furiously", (Chomsky, 1957/2002, p. 15) which sentence also contains contradictions: colorless vs green. From beginning of development of modern logic, logicians, e.g., Wittgenstein (1922), have acknowledged that there are next to true (T) and false (F), 'meaningless', 'senseless', 'nonsense' sentences, which adjectives may amount to more than two truth-values, one or more extra truth-value.

According to Wittgenstein philosophical sentences are senseless (#4.003 – '…. Most propositions and questions, that have been written about philosophical matters, are not false, but senseless' [contrary to propositions of science], pseudo-propositions are senseless (#4.1272 '….senseless pseudo-propositions ….'), laws of reference are senseless (#5.132 '…. Laws of inference, which-- as in Frege and Russell--are to justify the conclusions, are senseless and would be superfluous') and, famously, #6.54 'My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless …').

'Meaningless' (Carnap, 1967), 'senseless' and 'nonsense' may be considered third truth-values (if not 'meaningless', 'senseless' third and 'nonsense', 'nonsensical' fourth truth-value etc.), similar to 'undefined', although the *Liar*-type sentences may be considered false because they do not exist in external reality as from a Wittgensteinian point of view (see also Kripke 'ungrounded').[5]

B. Truth in fiction (reference in fiction).

C. Tenses (temporal logic).

D. Metaphors (contextualism[6]), some say metaphors belong to another language game, they can't be fully translated into plain, literal language, some say sense and truth of metaphors are context dependent.

> Finally, the question whether metaphors can be 'true', or rather 'fitting', 'correct' etc. According to Black [1954] metaphors belong to <u>another language game</u> than fact stating sentences, and as soon as there can be spoken of truth or falsehood there is no metaphor anymore, but literal, conventional use. Bartsch [1987] introduces a notion of <u>context dependent truth, semantical meaningfulness</u>. A sentence is semantical meaningful when the sentence is satisfied with respect to the referents of the referring constituents in that part of the discourse or situational setting. (Hertogh, 1989 – underline and bracketed remarks added)

E. More problems that have been resolved using alternate logics, that may be considered extensions to classical bivalent logic, such as deontic logic (logic of 'must',

ISSN – L 2805 – 1815

'should' etc.), doxastic logic (logic of 'to believe' etc.), epistemic logic (logic of 'to know' etc.), modal logic (e.g., box and diamond operator), PWS (possible world semantics, in which symbolism, the aforementioned alternate logics could possibly be expressed).
F. Interjections, hesitations, mistakes, slips of the tongue, intonations etc., if not considered only emotional.

Above enumerated problems are studied in linguistics, logic, research projects of formalization of natural language. In fact, there are no deep problems here, since, although one may call for revisions, one eventually only adds extensions.

Falsification of Logical Laws

If the *Liar* wants to deny logical laws like LEM and LNC, the Liar could be formalized as the very denial, that is, negation of universal validity of these laws.[7]

With help of (metalogical) proposition logic, predicate logic and truth predicate Tx (x is True)

¬LEM [1]

$\neg(P \vee \neg P)$

$\neg\forall x\,(Tx \vee \neg Tx)$

$\exists x\,(Tx \vee \neg Tx)$

¬LNC [2]

$\neg\neg\,(P \wedge \neg P)$

$\neg\forall x\,\neg(Tx \wedge \neg Tx)$

$\exists x\,\neg\,(Tx \wedge \neg Tx)$

Communicative Account Proposal

About the semantic versions of the *Liar*, we propose a pragmatic Gricean account, analyzing Epimenides's version of the *Liar* as falsification of generic tacit conversational principle—violation of Grice (1975) maxim of quality and maxim of manner--that people usually speak truth and perspicuously. By application of TE Matrix Epimenides-style paradox is extended to a valid and sound modus tollens argument (Popper, 1935, 1959/2002) with logical force of enthymeme (Aristotle, 1924). Maxims of quality and manner are derived from Grice Cooperative Principle (Grice, 1975, pp. 45-47):

Cooperative Principle (CP): Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Maxim of QUANTITY: quantity of information:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required

Maxim of QUALITY: Try to make your contribution one that is true:

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Maxim of RELATION, relevance: Be relevant.

Maxim of MANNER: Be perspicuous.

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

Logical Analyses

Maxim of quality, people usually speak truth[8], and maxim of manner, people usually speak perspicuously, are not just moral rules but empirical rules—if not, communication will be hampered. We can't deny (effective) communication is well possible, it runs the world, so, we may assume people indeed usually speak truthful and perspicuously.

We may apply (part of) apparatus to formalize natural language as with help of predicate logic.

Suppose

Cx	x is a Cretan
Sx	x Speaks (truth)
Lx	x Lies (do not speak truth)
X	person(s), human(s) (*variable*)
a	a definite, specific person, e.g., Epimenides (*constant*)
∀x	universal quantifier, *for all x it holds that...*

$\exists x$	existential quantifier, *for some, at least one, x it holds that …*
$\exists! x$	unique existential quantifier, *for exactly one x it holds that*
$\wedge$	conjunction
$\rightarrow$	(material) implication
$\neg$	negation

There is exactly one Cretan who says 'All Cretans are liars'.[9]

Which is the raw still invalid and unsound TE, bracketed in TE Matrix methodology

$$\left\lfloor \begin{array}{l} \exists! x\ Cx \wedge Sx\ (Ax\ Cx \rightarrow Lx) \\ Ca \wedge Sa\ (Ax\ Cx \rightarrow Lx) \end{array} \right\rfloor \qquad [2]$$

TE analyses may yield next valid and sound argument

P_0	$Ax\ (Cx \rightarrow Sx)$	[3]
P_1	$Ca \wedge Sa\ (Ax\ (Cx \rightarrow Lx))$	
P_2	$\exists! x\ Cx \wedge Sx\ (Ax\ (Cx \rightarrow Lx))$	
P_3	$Lx \leftrightarrow \neg Sx$	
P_4	$\exists! x\ Cx \wedge Sx\ (Ax\ (Cx \rightarrow \neg Sx))$	
P_5/C	$\neg(Ax\ (Cx \rightarrow Sx))$	

Or shortly:

$P'_0\ \forall x\ (Cx \rightarrow Sx)$ Grice´s maxim of quality P'_1

$Ca \wedge Sa\ (Ax\ (Cx \rightarrow \neg Sx))$ Epimenides *Liar* paradox

$P'_2/C\quad \neg\forall x\ (Cx \rightarrow Sx)$

It is about falsification of generic tacit conversational principle –Grice (1975) maxim of quality– that people usually speak truth, amounting to a valid and sound modus tollens argument (e.g. Popper, 1935, 1959/2002) with logical force of enthymeme (e.g. Aristotle,1924); falsification of generic rules by counterexamples (e.g. paradoxes) may be formalized as modus tollens (according to Popper, 1935, 1959/2002), and formula [4] may not exactly match the structure of modus tollens (P → Q, ¬Q, therefore ¬P), but it is reducible to it.

We could add it is not (only) about violating of maxim of quality, but (also) about violating of maxim of manner (e.g. on antinomic interpretations of the paradox)—apart from false, the *Liar*-type statements may (also) be considered paradoxical, that is in terms of the 4th maxim – obscure and ambiguous.

Also Eubelids' version, 'This statement is false' [d], interpreted as an antinomy leads to conflicting inferences (if *d* is true, then *d* is false; if *d* is false, then *d* is true – see note 2), which is a kind of strong ambiguity and obscurity, if not contradiction, and therefore it violates maxim of manner, defect of clarity, perspicuity, confusing of zero and metalevel; and one may as well say that it is false, violating maxim of quality.

Interpretations (Conclusions)

1. Although paradoxes and antinomies like *Liar*-type statements don't live up to all conditions of TE definition (they particularly fail the pragmaticist condition), they may be analyzed and interpreted with help of TE methodology of TE Matrix (logically analyzed as Extended Arguments) and TE Diagram (surveys).
2. The *Liar*, e.g., on Tarski's orthodox account, is a paradigmatic mathematical strong thought experiment, that as a paradox may be considered a metalogical instrument that questions its very presuppositions or the presuppositions of the logical system it is part of. We may particularly discuss semantic (instead of set theoretical) versions, such as Epimenides' (7th-6th cent. BCE) and Eubulides' (4th cent. BCE).
3. About discussions about classical logic and nonclassical logic as three-valued logics in attempts to resolve *Liar*-type paradoxes. We show that there are already more than two 'truth-values' used in the logical systems of Rudolf Carnap, Bertrand Russell, Ludwig Wittgenstein *Tractatus*—next to T(rue) and F(alse), respectively, 'meaningless'; nonsense, just a noise without meaning; 'nonsense' ('unsinnig'), 'senseless' ('sinnlos'). For the moment we may conclude that laws of logic, such as LEM and LNC, may have lost their categorical, apodictic status.
4. We propose a Gricean account of breaching of communicative norms, analyzing Epimenides's version of the *Liar* (The Cretan says, 'All Cretans are liars') as

falsification of generic tacit conversational principles –Grice (1975) maxim of quality and maxim of manner– that people usually speak truthful and perspicuously. This analysis may also hold for Eubulides-type *Liar* paradoxes as 'this sentence is false', since the very statement violates Grice's maxims of manner and quality, it is obscure, ambiguous (for directly or indirectly paradoxical), if not false. Thus, by application of TE Matrix, Epimenides-style and Eubulides-style paradoxes can be extended to valid and sound modus tollens arguments (Popper, 1935, 1959/2022) with logical force of enthymeme (Aristotle, 1924).

5. Historical interpretations of most renowned modern logicians show a variety of views on the *Liar*- type paradoxes. Russell considers the set theoretic *Russell's Paradox* and semantic *Barber Paradox* nonsense—"... the whole question whether a class is or is not a member of itself is nonsense ... because the whole form of words is just a noise without meaning." (Russell, 1919, p. 355). In *Tractatus* #3.24 Wittgenstein may have shown a classical logical resolution of *Liar*-type paradoxes, holding that propositions about complexes, which do not exist, are 'not nonsense but simply false.' Kripke (1975) nonclassical view involving (Kleene's) three truth-values ('ungrounded' paradox, so no truth-value applicable, but "undefined") may be interpreted within a classical metalogical framework on at least an "alternate intuition" as Kripke himself contends—"The term 'three-valued logic', occasionally used here, should not mislead. All our considerations can be formalized in a classical metalanguage" (Kripke, 1975, p. 701, note, 18). Textual analyses, semiotics, and speech act theory add a plethora of examples of paradoxical self- reference focusing on analyses of 'lying'. Neoclassical accounts, accepting paradoxes as 'brute facts' (Priest, 1979) may develop paracomplete and paraconsistent analyses assuming breaching of resp. LEM and LNC. Since there is no *communis opinio* the *Liar* may remain a conundrum. As we may have shown by a Gricean account a two-valued classical logical resolution is possible but LEM and LNC may have lost their status of categorical universal truth. In terms of Sorensen (1992) the *Liar* could be considered a *necessity refuter TE*, and after Tarski (1933/1935,1969, 1983a, 1983b) a *crucial consistency test TE*. It could be visualized by Necker cube (Wittgenstein *perceptual ambiguity*), or a Penrose tribar (or triangle) (*perceptual impossibility*), what Hofstadter could call a *feedback loop*,

related to core of consciousness (Hofstadter, 1979/1999, 2007), possibly rather the beginning than the end of human formalization attempts.

6. TE Diagram survey during and after conference Perspectives of Truth 2 in Bucharest (RO/EU), September 29, 2023, shows that *LEM* scores far lower (0.6) than *Pythagorean Theorem* (1.0) and *Einstein's Relativity* (0.9), possibly indicating that there is less academic confidence in classical logic nowadays (results of TE Diagram are not statistically significant since n =10). See Hertogh (2022) for more explanation and examples of TE Diagram surveys.
7. Possibly a rather farfetched speculative implication. We have followed extension option in this paper but on a strong interpretation of *Liar*-type sentences, they may show that there could be a mismatch between axiomatized mathematical systems and human and nonhuman reality of Planet Earth, and fundamental revisions may be needed. For this reason, among more reasons, we advocate a *Descriptive Semantics View*, instead of an axiomatized prescriptive system, without a forcing valuation function (still extension or methodological modesty than revision). For this reason, one could try and replace mathematics with another foundational discipline and science, e.g., linguistics (Wittgenstein *Philosophical Investigations*), biology (evolutionary epistemology), and ecology (end of Anthropocene). One may even argue that there is also a mismatch between mathematically based natural science and human and nonhuman reality of Planet Earth, which may eventually result in mismatches between natural science theory and reality, like we suffer now in the climate crisis, resulting from irresponsible technological societal applications of natural science theories, that have already been proven many times to be harmful to (wo)man and nature, human health and global ecology (environmental pragmaticist 21st century condition of semantics of thought experiments).
8. We may express global cross-cultural and environmental pragmaticist concerns by an addition to Popper's formula of progress of science (Popper, 1979, p. 243), thereby modifying it into *Progress of Science and Society View*

P_1	→	TS	→	EE/EP GC	→	P_2
Problem situation 1	→	Tentative Solutions	→	Error Elimination	→	Problem situation 2

In the 21st century Error Elimination (EE) could focus on *Environmental Pragmaticism* (EP) (countering pollutive effects of technological societal applications of science) and *Global Cross-culturalism* (GC) (theories should hold for, say, at least two cross-cultural communities). (Hertogh, 2015, 2018, 2020)

Acknowledgements

The research into antinomies of the *Liar* has benefited from academic presentations at two conferences in 2023, *Perspectives About Truth 2*, organized by Politehnica University of Bucharest (RO/EU), September 29-30, 2023, special thanks to Paula Tomi, and *Workshop New Perspectives on Anomalies in the Sciences*, organized by Universidade Federal do Rio de Janeiro (BR), October 5-6, 2023, special thanks to María del Rosario Martínez-Ordaz).

Notes

[1] Or, gender neutral, does a hairdresser who dresses hair of all those, and those only, who do not dress their hair themselves, dress her/his hair?

[2] There are many types and taxonomies of paradoxes, e.g., *seeming* or *apparent paradoxes*, that may be resolved on analyses, direct paradoxes (*contradictions*) and indirect paradoxes (*antinomies*). *Liar*-type sentences are often antinomies, where there is a conflict or contradiction between inferences drawn from a paradoxical sentence, e.g., Eubelids'-style

This statement is false. [d]

Inference 1:
Suppose [d] is true, then it is, in fact, false (contradiction)
Inference 2:
Suppose [d] is false, then it is, in fact, true (contradiction)

In accordance with definition in Paradoxes section, it is about intermingling of zero and meta level, possibly similar to what Hofstadter (1979/1999) calls *feedback loops*, possibly similar to *kōans*, visualized by optical illusions and impossible objects, such as Necker cube (see note 5, *Tractatus* #5.5423) and Penrose tribar etc., of which Penrose says 'It is clear that the "3-dimensional object" which the drawing [of a Penrose tribar] apparently depicts cannot exist in ordinary Euclidean space' (Penrose, 2004, p. 992) and '(An impossible object is a drawing of a solid figure that cannot exist because it embodies self-contradictory elements)' (Penrose, 1989, p. xv). See also note 5, Wittgenstein on Necker cube), and visit https://ia801208.us.archive.org/6/items/RoadToRealityRobertPenrose/road%20to%20reality-robert%20penrose.pdf PDF version of Penrose, 2004, *The Road to Reality,* which shows Penrose Tribar, as an example of an 'impossible object', to top of p. 992, fig. 33.21, Ch. 33, §33.9. (See also Hertogh, 2018, p, 269, note 7; Hertogh, 2021b, on *Huàtóus* and *Kōans*.) Hertogh (2025), on Dao De Jing, p. 54, figure 4, Penrose triangle – Source: https://en.wikipedia.org/wiki/List_of_optical_illusions

[3] Kripke seems to be unsure about the truth-gap approach and natural language

> On the basis of the fact that the goal of a universal language [e.g., Tarski] seems elusive, some have concluded that truth-gap approaches, or any approaches that attempt to come closer to natural language than does the orthodox approach, are fruitless. I hope that the fertility of the present approach, and its agreement with intuitions about natural language in a large number of instances, cast doubt upon such negative attitudes. (Kripke, 1975, p. 715)
>
> I am somewhat uncertain whether there is a definite factual question as to whether natural language handles truth- value gaps- at least those arising in connection with the semantic paradoxes. (Kripke, 1975, p. 712)

[4]Ripley:

> This paper provides a defense of the full strength of classical logic, in a certain form, against those who would appeal to semantic paradox or vagueness in an argument for a weaker logic.
>
> I will not argue that these paradoxes are based on mistaken principles; the approach I recommend will extend a familiar formulation of classical logic by including a fully transparent truth predicate and fully tolerant vague predicates. It has been claimed that these principles are not compatible with classical logic; I will argue, by both drawing on previous work […] and presenting new work in the same vein, that this is not so. We can combine classical logic with these intuitive principles, so long as we allow the result to be nontransitive. In the end, I hope the paper will help us to handle familiar paradoxes within classical logic; along the way, I hope to shed some light on what classical logic might be for. (2013 , p. 1, underline added)

[5] Wittgenstein also mentions theory of probability with many truth-values, example of many- valued logic "4.464 The truth of tautology is certain of propositions possible, of contradiction impossible. (Certain, possible, impossible: here we have an indication of that gradation which we need in the theory of probability.) (Wittgenstein, 1961, p. 54)

In Russell's Wittgenstein interpretation sentences about ethics are 'mystical' or 'inexpressible' (possibly again additional truth-values)--see Wittgenstein (1922, p. 18) from introduction by Russell: "…. The whole subject of ethics, for example, is placed by Mr Wittgenstein in the mystical, inexpressible region...". However, we could possibly assume antinomies as the *Liar* may be considered complex and non-existent and are therefore not nonsense but false (similar to Kripke's 'alternate intuition'). Every statement about complexes can be analysed into statements about their constituent parts (#2.0201); in case of the *Liar*, the proposition appears to be complex because of its reference that is equivocal, ambiguous--non-existent (#3.24, similar to Kripke's 'ungrounded'), possibly similar to Necker Cube perceptual ambiguity (#5.5423)

> 3.24 A complex can only be given by its description, and this will either be right or wrong. The proposition in which there is mention of a complex, if this does not exist, becomes not nonsense but simply false.... (Wittgenstein, 1922, p. 33)
>
> 5.5423 To perceive a complex means to perceive that its constituents are combined in such and such a way. This perhaps explains that the figure
>
> [image Necker cube with letters added]
>
> can be seen in two ways as a cube; and all similar phenomena. For we really see two different facts. (If I fix my eyes first on the [lower] corners *a* and only glance at [higher] *b*, *a* appears in front and *b* behind, and vice versa.) (Wittgenstein, 1922, pp. 71-72, bracketed remarks added)

Visit https://www.gutenberg.org/files/5740/5740-pdf.pdf Gutenberg version of *Tractatus*, which shows Necker cube, and letters added by Wittgenstein, on p. 72 and German text on p. 142; see also Hertogh (2025), p. 54, figure 3, Necker cube--without any numbers--Source: https://en.wikipedia.org/wiki/List_of_optical_illusions).

Daniel Rowe (Oxford University) remarked during conference Perspectives about Truth 2 that Wittgenstein does not think the *Liar* is a philosophical important problem, but that Kripke does. For more Kripkean takes of the *Liar*, see Rowe, 2023, 'Temporal or Staged Logic and the Disambiguation of Semantic Paradoxes.' Remarks on use of 'senseless' and 'nonsense' in *Tractatus* are based on Ogden/Ramsey translation (Wittgenstein,1922 – preferring 'senseless') and could be double-checked by more translations (as Pears/McGuinness, Wittgenstein, 2022 – preferring 'nonsensical') and German original (Wittgenstein, 1922, 2022 -- 'sinnlos' and more often 'unsinnig'), but it may result in same conclusion viz. that there are used more than two truth-values in *Tractatus*.

[6] We don't think contextualism may entail relativism, rather relativity, saying that different parts, domains of reality are formalized in a different way. In analyses of metaphors (Hertogh, 1989) we distinguish between standard semantic metaphors and context-dependent or pragmatic metaphors. The latter may be identified, analysed and formalized with help of thematic dimensions, a semantic term for context; in case of metaphors, it is most often about the actual context, that is the sentence in which the metaphor theme appears. Different parts of reality (domain, discourse etc.) may be analysed and interpreted different as in accordance with possibly different structure, texture, substrate etc. of different parts of reality, logic may not only study declarative sentences (Wittgenstein, 1922, Russell & Whitehead, 1910-1913; Carnap, 1928, 1967), but many more moods, such as questions, irony etc. (Wittgenstein, 1953; Searle, 1969; Grice, 1975, etc.). E.g. Galileo's gravity theory may hold (by approximation) for Planet Earth,

Einstein's relativity theory for micro- and macrocosm. Voodoo and Hollywood Zombies should be analysed different from *p*- or philosophical *Zombies* (see Chalmers, 2017; Hertogh, 2023c). Poetical metaphors may be analysed different from philosophical metaphors, the former may be close to similes, the latter are often a tip of a (theoretical) submerged model. It would be a kind of word magic to suggest that a word may mean the same in each context. Historical and etymological research may show how word meanings develop by metaphorical and metonymical extensions, whereby substantial use of a word in a new context may result in a new (dictionary) sense.

[7] LEM, LNC go back on Aristotle, *On Interpretation*, *Metaphysics* (Aristotle, 1908-1952). These basic logical laws appear in more cultures, e.g., in Mohist logic as 'basic principles regulating disputations' (see Zhang & Liu, 2007).

[8] Martinich compares Grice's maxim of quality to Searle's sincerity condition (Martinich, 1980, p. 226). Fallis tries to define and categorize instances of lying (Grice--'Do not say what you believe to be false') in terms of intentions, beliefs, misspeakings, sarcasms, unawareness, insincere assertions, and concludes it is about violating a norm: "Even so, all liars, including young children, do intend to communicate something false by saying that thing. Thus, they do intend to do something that would violate the norm against communicating something false if it were in effect". (Fallis, 2012)

It takes half a century from Richard Nixon's exceptional Watergate scandal lying (discussed in Kripke, 1975) to David Trump's allegedly habitual lying in 2010s, 2020s. One may wonder, isn't the present-day USA President breaching Grice's maxims? And an answer could be that the fact that the President is allegedly lying, is communicated to him and exposed in media and academics (Hoinarescu, 2018) continuously, called fact checking, which may prove Grice's maxims are still alive today, and that there have been introduced terms like misinformation and disinformation—next to President's disparagement of the media as fake news--to handle a situation of possibly increasing false, misleading and biased information in communication of governments, companies, social media etc., whether or not one may reduce these discussions to community debates only (e.g., Facebook). (See Broda & Strömbäck, 2024, concluding 'Considering the threat misinformation, disinformation, and fake news pose, it is vitally important that we … continue advancing the field. The stakes are undoubtedly high (p. 161).)

Critical thinking could help deciding whether a claim is always true, sometimes true, partly true, or false (see Hertogh, 2015, p. 64, section on logic and critical thinking, pp. 60-74; see also Hertogh, 2022).

[9] With regard to possibly uniquely defining existential quantification– It is more likely that Epimenides (or any other one Cretan) is lying than that all Cretans are lying, isn't it? (as depending on your presupposition whether or not you think people usually speak truth, or are usually lying, which is in fact exemplified with regard to Cretans in P_0/P'_0).

Falsification –so, it is not surprising that according to Gödel and Tarski these exceptions may falsify parts of arithmetic, logic, mathematics as a consistent and coherent system.

References

Anapolitanos, D. A. (1991). 'Thought Experiments and Conceivability Conditions.' In T. Horowitz, & G. Massey (Eds.) *Thought Experiments in Science and Philosophy*. Rowman and Littlefield (pp. 87-98). https://philsci-archive.pitt.edu/3190/1/thought_experiments_toc.htm

Aristotle. (1908-1952). *The Works of Aristotle, Translated into English Under the Editorship of W.D. Ross, 12 Vols.* Clarendon Press.

Aristotle. (1924). *Rhetorica, The Works of Aristotle, Vol. 11* (Trans. Roberts, W. Rhys). Clarendon Press.

Asenjo González, F. (1966). A Calculus of Antinomies. *Notre Dame Journal of Formal Logic*, 7(1), 103-105.

Barthes, R. (1973). Analyse textuelle d'un conte d'Edgar Poe. In C. Charbrol, *Sémiotique* *narrative et textuelle* (pp. 29-54). Librairie Larousse.

Barthes, R. (2014). The Death of the Author Textual Analysis: Poe's "Valdemar".' (Trans. G. Bennington). In D. Lodge, & N. Wood (Eds.) *Modern Criticism and Theory: A Reader* (pp. 311-336, Ch. 17). Third Ed. Routledge.

Bartsch, R. I. (1987). *Norms of Language, Theoretical and Practical Aspects.* Longman.

Beall, Jc., Glanzberg, M., & Ripley, D. (2016/2023). Liar Paradox. *Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (Eds.).

Black, M. (1954). Metaphor. *Proceedings of the Aristotelian Society*, 55, 273-294.

Broda, E., & Strömbäck, J. (2024). Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review. *Annals of the International Communication Association*, 48.(2), 139–166. https://doi.org/10.1080/23808985.2024.2323736

Carnap, R. (1928). *Der logische Aufbau der Welt*. Felix Meiner Verlag.

Carnap, R. (1967). *The Logical Structure of the World Pseudoproblems in Philosophy*. (Trans. R. A. George). University of California Press.

Chalmers, D. J. (2017). *Zombies on the Net.* https://consc.net/zombies-on-the-web/

Chomsky, N. (2002). *Syntactic Structures, Second Edition with an Introduction by David W Lightfoot.* Mouton de Gruyter. (Original work published 1957)

Da Costa, N. C. A. (1974). On the Theory of Inconsistent Formal Systems. *Notre Dame Journal of Formal Logic*, 15(4), 497–510.

Einstein, A., & Infeld, L. (1938). *The Evolution of Physics-From Early Concept to Relativity and Quanta*. Cambridge University Press.

Epstein, M. (2015). 'Paradoxical Speech Acts: Transformatives and Counterformatives.' *Parallax*, 21 (2), 134-142.

Fallis, D. (2012). Lying as a Violation of Grice's First Maxim of Quality. *Dialectica*, 66(4), 563-581.

Galileo G. (1638/1914). *Mathematical Discourses and Demonstrations, Relating to Two New Sciences/Dialogues concerning Two New Sciences*, Translated from the Italian and Latin into English by Henry Crew and Alfonso de Salvio. With an Introduction by Antonio Favaro (Trans. Galileo 1638. *Discorsi e Dimostrazioni Matematiche, Intorno à Due Nuove Scienze.* Elsevirii). MacMillan.

Gödel, K. (1931). Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38, 173–98.

Gödel, K. (2000). *On Formally Undecidable Propositions of Principia Mathematica and Related Systems.* Translated by B. Meltzer, Introduction by R.B. Braithwaite. Dover Publications.

Grice, H. Paul (1975). Logic and Conversation. In P. Cole, & J. Morgan (Eds.) *Syntax and Semantics Vol.3 Speech Acts* (pp. 41-58). Academic Press.

Hertogh, C. P. (1989). Identification, Analysis, and Interpretation of Metaphorical Indicative Sentences of Subject-Predicate Form. Trans. Conclusion Unpublished Master's Thesis University of Amsterdam.

Hertogh, C.P. (2015). Semantics of Thought Experiments. PhD diss. VUB Brussels.

Hertogh, C.P. (2018). Logical Analyses of Vipassanā Meditation. *Contemporary Buddhism*, 19 (2), 251-271.

Hertogh, C.P. (2020). Goodman's Grue—Relativized Pluralism and Paradigmatic Thought Experiment (Whack). *Humanities Bulletin*, 3(1), 64-103.

Hertogh, C.P. (2021a). Computation and Visualization Thought Experiments after Lakatos's Heuristic Guessing Method. In M. Trepczyński (Ed.) *Philosophical Approaches to the Foundations of Logic and Mathematics in Honor of Professor Stanisław Krajewski* (pp, 271-297, Ch. 10). Leiden.

Hertogh, C.P. (2021b). Thought Experiment Analyses of Huàtóus and Kōans. In Rivas, V. (Ed.), (2021). *PAP Transactions* (pp. 93-107). 2nd USAPP Summit, July 21-24.

Hertogh, C.P. (2023a). Thought Experiment Analyses of the Liar. Presentation Perspectives about Truth 2, September 29-30, 2023, Bucharest.

Hertogh, C.P. (2023b). Semantics of the Liar. Presentation Workshop New Perspectives on Anomalies in the Sciences, October 5-6, 29-30, 2023, Rio de Janeiro.

Hertogh, C.P. (2023c). The Conceivability of Consciousness--Kirk's and Chalmers's Zombie Thought Experiments. *London Journal of Research in Computer Science and Technology*, 23(1), 21-41.

Hertogh, C.P. (2024). Thought Experiment Analyses of Albert Einstein's Elevator [Análisis del experimento mental del elevador de Albert Einstein]. *Analítica* (4), 10-40.

Hertogh, C.P. (2025). Predication and Opposition in Dào Dé Jīng Thought Experimental Analyses of 道可道非常道(III)– Global Cross-cultural and Religious Aspects, *Humanities Bulletin*, 8(1), 20-56.

Hofstadter, Douglas R. (1999). *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books. (Original work published 1979)

Hofstadter, D. R. (2007). *I Am a Strange Loop*. Basic Books.

Hoinarescu, L. (2018). Lying as Speech Act. A Sociodiscursive and Interactional Approach. *Revue Roumaine de Linguistique*, 63 (1–2), 147–166.

Horowitz, T. & Massey, G. J. (Eds.). (1991). *Thought Experiments in Science and Philosophy*. Rowman and Littlefield. https://philsci-archive.pitt.edu/3190/1/thought_experiments_toc.htm

Kearns, J. (2007). An Illocutionary Logical Explanation of the Liar Paradox. *History and Philosophy of Logic*, 28, 31–66.

Kirk, G.S., Raven, J. E., & M. Schofield (1983). *The Presocratic Philosophers: A Critical History with a Selection of Texts* (Second Edition). Cambridge University Press.

Kleene, S. C. (1952). *Introduction to Metamathematics.* Van Nostrand.

Kripke, S. A. (1975). Outline of a Theory of Truth. *Journal of Philosophy*, 72, 690–716. Reprinted in Martin, R. L. (Ed.). (1984). *Recent Essays on Truth and the Liar Paradox* (pp. 54-81). Clarendon.

Kripke, S.A. (1980). *Naming and Necessity*. Basil Blackwell.

Kripke, S.A. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.

Kripke, S.A. (2019). 'Naming and Necessity Revisited.' Center for Logic and Language Special Lecture May 30, 2019, Institute of Philosophy, University of London. [Video] https://www.youtube.com/watch?v=3zazonG6zBk

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Enlarged Second Ed. (First Ed. 1962) University of Chicago Press.

Martinich, A. P. (1980). Conversational Maxims and Some Philosophical Problems. *The Philosophical Quarterly*, 30(120), 215-228.

Martin, R. L. (Ed.). (1970). *The Paradox of the Liar*. Yale University Press.

Martin, R.L. (Ed.). (1984). *Recent Essays on Truth and the Liar Paradox.* Clarendon Press.

Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics.* Oxford University Press.

Penrose, R. (2004). *The Road to Reality, A Complete Guide to the Laws of the Universe*. Jonathan Cape: London. https://ia801208.us.archive.org/6/items/RoadToRealityRobertPenrose/road%20to%20reality- robert%20penrose.pdf

Popper, K. R. (1935). *Logik der Forschung.* Julius Springer Verlag.

Popper, K.R. (2002). *The Logic of Scientific Discovery*, Translation by the Author of *Logik der Forschung* 1935. Routledge Classics. (Original work published 1959)

Popper, K.R. (1979). *Objective Knowledge: An Evolutionary Approach, Revised Edition*. (First Ed. 1972) Clarendon Press.

Priest, G. (1979). Logic of Paradox. *Journal of Philosophical Logic*, 8, 219-241.

Priest, G. (1984). Logic of Paradox Revisited. *Journal of Philosophical Logic*, 13(2), 153–179.

Quine, W.V.O. (1976). *Ways of Paradox*. Harvard University Press.

Ripley, D. (2013). Revising up: Strengthening Classical Logic in the Face of Paradox. *Philosophers' Imprint*, 13(5), 1-13.

Rowe, D. (2023). Temporal (or Staged) Logic and the Disambiguation of Semantic Paradoxes. *Presentation Perspectives about Truth 2*, September 29-30, 2023, Bucharest.

Russell, B. (1919). The Philosophy of Logical Atomism, VII The Theory of Types and Symbolism: Classes, VIII Excursus into Metaphysics: What There Is. *The Monist*, 29 (3), 345–380.

Russell, B. & Whitehead, A. N. (1910–1913). *Principia Mathematica*. 3 Vols. Cambridge University Press.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Sorensen, R. A. (1992). *Thought Experiments.* Oxford University Press. http://course.sdu.edu.cn/G2S/eWebEditor/uploadfile/20121224164821178.pdf

Sutrop, U. (2009). Wittgenstein's Tractatus 3.333 and Russell's Paradox. *Trames, Journal of the Humanities and Social Sciences*, 13(2), 179–197. https://doi.org/10.3176/tr.2009.2.06

Tarski, A. (1969). Truth and Proof. *Scientific American*, 220 : 63–77 (63-70, 75-77).

Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen (L. Blaustein, transl., with a postscript added), *Studia Philosophica*, 1, 261–405. [Original work published 1933]

Tarski, A. (1944). The Semantic Conception of Truth: And the Foundations of Semantics. *Philosophy and Phenomenological Research*, 4(3), 341–375.

Tarski, A. (1983a). *Logic, Semantics, Metamathematics.* (Second Ed.) (J. Corcoran, Ed.). Hackett.

Tarski, A. (1983b). The Concept of Truth in Formalized Languages. (J. H. Woodger, Transl.), in A. Tarski, *Logic, Semantics, Metamathematics* (Second Ed., pp. 152-278).). Hackett.

Wittgenstein, Ludwig (1922). *Tractatus Logico-Philosophicus.* With an Introduction by Bertrand Russell. Trans. C. K. Ogden. (Bilingual Ed.) Kegan Paul Trench Trubner. https://www.gutenberg.org/files/5740/5740-pdf.pdf [eBook #5740]

Wittgenstein, L. (1953). *Philosophical Investigations.* Basil Blackwell.

Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. Trans. McGuinness, B.F. and D.F. Pears, and with Introduction by Bertrand Russell. The Humanities Press.

Wittgenstein, L. (2022). *Tractatus Logico-Philosophicus Logisch-philosophische Abhandlung by Ludwig Wittgenstein* (First published by Kegan Paul London, 1922) *Side-by-side-by-side Edition, Version 0.63 (July 14, 2022) containing the original German, alongside both the Ogden/Ramsey, and Pears/McGuinness English translations.* https://people.umass.edu/klement/tlp/

Wójtowicz, A. (2021). An Analysis of Paradoxes. In M. Trepczyński (Ed.) *Philosophical Approaches to the Foundations of Logic and Mathematics in Honor of Professor Stanisław Krajewski* (pp, 258-270). Leiden.

Zhang, J. & , Liu, F. (2007). Some Thoughts on Mohist Logic. In J., van Benthem, S. Ju, & F. Veltman (Eds.)(2007). *A Meeting of the Minds: Proceedings of the Workshop on Logic, Rationality and Interaction*. College Publications (pp. 85-102) . https://www.researchgate.net/publication/240798715_Some_Thoughts_on_Mohist_Logic

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 120-135

Artículo/Article

Radical contextualism and open compositionality: a metatheoretical proposal

Contextualismo radical y composicionalidad abierta: una propuesta metateórica

Pablo David Chávez Carvajal

Universidad Nacional Autónoma de México, México
pablodavidcc@gmail.com
https://orcid.org/0000-0002-1457-4705

DOI https://doi.org/10.48204/2805-1815.8476

INFORMACIÓN DEL ARTÍCULO

Recibido el: 15/7/2025

Aceptado el: 31/8/2025

Keywords:

Philosophy of language, semantics, radical contextualism, compositionality, open compositionality

Palabras clave:

Filosofía del lenguaje, semántica, contextualismo, composicionalidad, composicionalidad abierta

ABSTRACT/RESUMEN

Abstract:

In this paper, I explore the relationship between radical contextualism and compositionality. Radical contextualism is a family of theories defending that a sentence's meaning in natural language is always invariant because it is too broad and unspecified to be otherwise and requires contextual adjustment. Compositionality is the idea that a sentence's complex meaning comprises its parts. I shall evaluate whether radical contextualism benefits from including compositionality to explain linguistic meaning in natural language. Including compositionality might allow us to formalize aspects of radical contextualism and explain meaning-formation more precisely. I shall argue, nonetheless, that the classical notion of composition fails to account for the interpretation of sentence meaning in natural languages (as guided by a radical contextualist theory). An open compositionality scheme is crucial since I understand meaning-determination as a decision-making task. To explain how we understand sentences in natural language, open compositionality must be used to formulate a radical contextualist theory. I shall also offer a novel metatheory encompassing open compositionality and radical contextualism.

Resumen:

En este artículo, exploro la relación entre el contextualismo radical y la composicionalidad. El contextualismo radical es un conjunto de teorías que defienden que el significado de una oración en el lenguaje natural es siempre invariante porque es demasiado amplio y poco específico como para ser de otra manera y requiere un ajuste contextual. La composicionalidad es la idea de que el significado complejo de una oración se compone de sus partes. Evaluaré si el contextualismo radical se beneficia de la inclusión de la composicionalidad para explicar el significado lingüístico en el lenguaje natural. La inclusión de la composicionalidad podría permitirnos formalizar aspectos del contextualismo radical y explicar la formación del significado con mayor precisión. No obstante, argumentaré que la noción clásica de composición no logra explicar la interpretación del significado de las oraciones en los lenguajes naturales (tal y como lo indica una teoría contextualista radical). Un esquema de composicionalidad abierta es crucial, ya que entiendo la determinación del significado como una tarea de toma de decisiones. Para explicar cómo entendemos las oraciones en el lenguaje natural, se debe utilizar la composicionalidad abierta para formular una teoría contextualista radical. También ofreceré una nueva metateoría que abarca la composicionalidad abierta y el contextualismo radical.

Introduction

Roughly stated, the principle of compositionality (commonly attributed to Frege)[1] states that the meaning of a complex expression is determined by its constituents: for example, the complex meaning of a sentence is determined by its words and their structure. Compositionality has been used to explain how the meaning of an expression in a natural language is *composed*. Modeling language users' linguistic competence (i.e., accounting for how they understand sentence meanings) putatively explains how language users can learn natural languages and, in turn, how they understand sentences they have not heard before. Several scholars have, however, argued that radical contextualism is largely incompatible with compositionality.

Radical contextualism in the philosophy of language is a family of theories that more or less follow Recanati (2010) in defending the following thesis: "[I]n general (i.e. not only in the special case of indexicals), the propositional contribution of an expression is not fully determined by the invariant meaning conventionally associated with the expression type but depends upon the context". (p. 17)

This view can be divided into moderate and radical contextualism.

- Moderate contextualism assumes that the proposition a sentence conveys is semantically complete and its truth value is determined by context-sensitive aspects of the sentence's use.

- In radical contextualism, the invariant meaning is always unspecified, too wide to be made specific, and in need of contextual adjustment. This suggests a high degree of contextual sensitivity.

I shall concentrate exclusively on radical contextualism. The difficulty in reconciling it with compositionality lies in (a) how grammar allows the assignment of a complex sentence's interpretation and (b) how language users understand complex sentences, guided only by their context (the time and place of utterance).

Consider the sentence:

We have a lame duck situation **[1],**

uttered in three different contexts:

1. *In a park*. It would then have a conventional meaning, namely that there is a lame duck and it probably needs help.
2. *In a business environment in the United States*. Here, the situation is that a manager is about to step down.
3. *In a cricket match*. A batter has been dismissed without making runs, and this has happened uninterestingly.

The first context is the only one that can be interpreted purely through grammar. The other two interpretations require contextual adjustment.

The following questions thus arise: can the notion of compositionality be preserved in a radical contextualist theory? Would compositionality's inclusion in radical contextualism be of any use? Do we need a notion of compositionality that differs from classical compositionality? To answer these questions, I shall first critically describe what compositionality is, some of its applications and variations, and the most common objections. I shall then discuss the relationship between compositionality and contextualism and then characterize contextualism in general to highlight the difficulties this relationship faces. Next, I shall formulate a possible radical-contextualist response—one that considers Recanati's (2010) proposal to weaken classical compositionality. However, such a reply fails because it does not resolve the salient variations that can appear when defending the contextualist thesis—mainly that a sentence's meaning always requires contextual adjustment.

However, classical compositionality suggests that the level of meaning which we should consider compositional is not radically context sensitive. According to classical compositionality, this level should be sufficient; to maintain radical contextualism motivated by contextual sensitivity, we should therefore adopt a theory of open compositionality, as García-Ramírez (2019) suggests. This allows for the inclusion of significant variability in linguistic meaning, along with compositionality. I shall explain this notion and argue that open compositionality is the only way for radical contextualists to preserve the advantages of compositionality. The discussion presupposes radical contextualism, but this is a family of theories that can differ from each other, so I shall present a suitable metatheory. This entails developing a set of propositions expressing the conditions a radical contextualist theory must fulfill to maintain compositionality.

What is compositionality and how useful has it been in the philosophy of language?

Language users can produce sentences they have never written or verbalized before. These sentences can, in turn, be understood by other language users who have not previously encountered them. For example, the following sentence (or one equivalent in meaning) has likely never been produced before:

A young rugby player and his uncle are playing bridge in a purple house in Montevideo, Minnesota. **[2]**

Most English speakers will have no difficulty in understanding this sentence, even if they have never heard or read it before, since they can identify the parts of the sentence. This indicates that the ability to understand a natural language's sentences lies in how their parts cohere to generate distinct meanings. The number of complex expressions language users can produce and understand seems to be unlimited. But how can limited beings like us be so linguistically productive? How do we learn any number of languages, and the unlimited number of meaningful expressions that can be formed from each language's more or less stable and restricted vocabulary? Such a phenomenon requires explanation, and the best hypothesis involves natural languages' compositionality. Here is how Elbourne (2011) characterizes the principle of compositionality: "The meaning of a complex phrase is determined solely by the meaning of its parts and their syntactic arrangement" (p. 28). (See Zimmerman, 2020, for a contemporary reconstruction of possible Fregean composition) This principle (or one of its variations) has been used to explain competent speakers' linguistic ability, which (hypothetically) allows such speakers of a natural language to understand an infinite set of linguistic expressions, based on their knowledge of specific, elementary categories.[2] This is called "productivity." As human beings with cognitive limitations, we cannot know an infinite number of meanings immediately. Elbourne explains that meaning is systematic because it allows complex expressions to be related. His principle thus establishes a methodological claim about how we should study meaning in natural language.

Nevertheless, Jaszczolt (2018) has noted the following: "Compositionality is not only a methodological claim but also an epistemological and metaphysical one, based on the argument of the productivity and systematicity of patterns of interaction in conversation". (p. 80)

Compositionality is epistemological because it lets us account for how speakers know meanings in natural language. This starts with their knowledge of languages' essential elements, which cannot be explained in more elementary terms. Szabó (2012) expresses the idea more precisely:

> The argument from systematicity states that anyone who understands a number of complex expressions *e*1..., *e*n understands all other complex expressions that can be built up from the constituents of *e*1..., *e*n using syntactic rules employed in building up their structures. Since this is so, there must be something competent speakers know (perhaps tacitly) based on which they can determine what the complex expressions built through such recombination mean. (p. 77)

Compositionality is metaphysical because it determines what kind of composition natural language sentences must have if they are to constitute meaningful expressions. An unordered juxtaposition of strings of characters implies a meaningless concatenation of marks and sounds. A truly meaningless strings of characters would be something like "Zdfbdf qrte szgwet" (i.e., completely random letters, assuming randomness is meaningless).

That said, there are concrete semantic theories that use compositionality. Examples include Montague's semantics (Dowty et al., 1981) and Davidson's (2001) truth-conditional semantics. (For a detailed study of this type of semantics, see Lepore &

Ludwig, 2009). However, these are semantic theories in the formal sense—theories that are shaped by the following properties:

> A compositional theory of meaning for a language L is a formal theory that enables anyone who understands the language in which the theory is formulated to understand the primitive expressions of L, and the complex expressions of L based on the understanding of the primitives. (Lepore & Ludwig, 2009, p. 18)

Considering the characteristics usually attributed to compositionality, we can add to this a fundamental property of natural languages: learnability. The arguments used to defend such notions are therefore referred to as *learnability arguments*. (This property is most notably defended by Davidson, 2001) Here is a typical one, in which L stands for any natural language (see Pagin, 2012, p. 514, where the argument comes from):

1. There is a sentence in L, in any context c, for each of the infinite number of existing meanings.
2. We can learn L (and thus understand all these sentences that convey each of the infinite number of meanings).
3. The mechanism of *directly learning* sentence meanings allows us to learn only a finite number of such meanings.
4. We can learn L, and understand expressions for the infinite number of meanings, only if we can follow compositional rules for constructing the infinite number of expressions in L, based on the finite number of expressions that can be directly learned (using the mechanism mentioned in Premise 3). In other words, we can learn L only if L has compositional semantics.
5. L has compositional semantics.

However, for Pagin (2012), the learnability argument[3] leads to a problem, which arises from the assumption that the infinity of sentences speakers never use are meaningful. This problem is generated by one of the argument's premises and a presupposition. In Premise 1, it is assumed that a natural language, such as English, has an infinity of meanings. At the same time, it is presupposed that humans cannot know an infinite number of meanings; rather, there are infinite meanings which humans can potentially learn. If so, English will have an infinite set of meanings that speakers do not use, leading to the production of linguistic meanings that seem unnecessary for communication. The best justification for holding that any natural language has infinite meanings is that language semantics is compositional: these unused sentences' meanings can be constructed, even though they are not used. However, if this is the justification for holding P1 (that there is an infinity of meanings we never use) then one commits a *petitio principii*: one presupposes compositionality rather than proving its necessity.

This demonstrates how undesirable consequences arise when we explain meaning composition as part of an attempt to formulate a compositional theory. We can see that "getting to the bottom of the composition of meaning is not a simple task. There are also difficulties involved in combining the meaning of adjectives and nouns" (Jaszczolt, 2023, p. 65). For instance, a falsified Picasso is, strictly speaking, not a Picasso. Since such difficulties arise when appealing to compositionality in natural language, one can ask: What results from contextual sensitivity's inclusion in the meaning of natural language sentences? By contextual sensitivity, I mean the following: *A sentence s will have the property of contextual sensitivity (cs) if and only if how s's meaning is understood changes depending on the occurrence of s's use, which is determined by the contextual characteristics (time and place) of s's use*.

Given how context can affect a sentence's meaning, speakers' context and intention when uttering a sentence can determine that meaning only broadly.

Consider sentence **[1]** again. Plausibly, we can determine which of its three meanings applies only when we recognize the speaker's communicative intention. It would, therefore, be challenging to find compatibility between contextual sensitivity and compositionality. My discussion will focus on the relationship between compositionality and radical contextualism—a thesis motivated by contextual sensitivity—because several scholars have argued that the two theses are incompatible.[4]

The theoretical incompatibility between compositionality and radical contextualism

How can meaning in natural language be studied if there is context sensitivity? How context determines sentence meaning is controversial. A context's representational accuracy depends on the contextualist theory postulating what role it plays or what it is. As Ciecierski & Grabarczyk (2020) have written:

> It is a truism that context, content, and meaning interact. On the one hand, context determines what is communicated in the discourse; on the other hand, what is communicated in the discourse shapes the situation in which the discourse is embedded. Philosophers who agree about these basic facts disagree about nearly everything else. (p. 1)

We can characterize context thus: "We call these things 'contexts' because they are being used to represent the concrete situation in which a sentence event takes place" (Stalnaker 2018, p. 105). We can define radical contextualism as follows:

> [In radical contextualism, the] invariant meaning of a non-indexical expression type is too unspecified and/or too rich to render a literal interpretation of its token in any context. Consequently, contextual adjustment is required in every context to determine its propositional contribution. (Huang, 2017, p. 964)[5]

However, if we require contextual adjustment then we also require more than an elementary notion of context (i.e., time and place). Plausibly, we can characterize context[6] in a very general way, compatible with my purposes here, to present a general radical-contextualist thesis. Pollock's characterization may be useful in this regard. She writes:

> Context here should be understood in the broad sense, to include things like interlocutors' background knowledge as well as physical features of the environment of the speech exchange. (Pollock, 2020, p. 249)[7]

As such, we can understand what is required to clarify the notion of radical contextualism as follows: Consider any natural language sentence S that requires contextual adjustment to fix its meaning M1 at the time of utterance. Such an adjustment will be achieved if and only if either the interlocutor's background knowledge or the physical characteristics of the environment in which S is used determine what M1 means at a particular time.

For example, as the Introduction explained, contextual adjustment is required for sentence **[1]**:

- This expression might be used in a business environment in the United States when a manager is about to step down. The interlocutor recognizes the meaning of the idiomatic expression "lame duck" (to express that this manager is about to leave), which represents a situation of interest to the people having the conversation.

- The expression might also be used in a cricket context when a batter has been dismissed without making runs and nothing makes this situation interesting. Recognizing the cricket jargon, the interlocutor knows that “duck” means that the batsman has not scored a single run. Given the circumstances of the match, however, this occurrence could have been more interesting.

We can therefore describe why radical contextualism is incompatible with compositionality if there must always be a contextual fit based on a sentence’s interpretation (when considering radical contextualism). How can we maintain the notion that the meaning of a complex sentence is determined solely by the meanings of its parts and their syntactic arrangement? Compositionality would hold only in the sentence’s grammar, so contextual adjustment could not require compositionality. This is because compositionality does not depend on factors that are non-linguistic, such as background knowledge and the environment of a sentence. Is there a way to maintain both radical compositionality and contextualism? One way involves postulating a series of rules that allow us to invoke the need for contextual adjustment when explaining the variability in a sentence’s linguistic meanings. As we have seen, this weakens compositionality.

To illustrate this, we can recognize the following context-sensitive lexical rule **(1*)**, as formulated by Recanati (2010):

$$I(a)c = f(c)$$

The characterization is as follows: the conventional meaning of a sentence *s* determines a function *f*, which, given a context *c*, momentarily alters its meaning (Recanati, 2010, 37).[8] More precisely, this rule models how a sentence’s contextual adjustment occurs (as discussed in Section 1). A sentence’s conventional meaning becomes a function of the context. In other words, the conventional meaning is adjusted to fit the context of the sentence. This provides a way to reconcile (i) what is grammatically determined in a sentence and (ii) the context sensitivity that can cause that sentence’s meaning to vary.

Let us return to **[1]** in its second meaning. The conventional meaning of “lame duck” appears as a contextually determined function in the business environment. However, other types of circumstances are excluded (e.g., that there is a mallard duck present or a cricket match going on). Having a theory that preserves compositionality (even while accepting that natural language displays contextual sensitivity) is highly desirable because Recanati's proposal produces rules that establish a procedure under which natural language sentences obtain various meanings.[9]

Briciu (2020) has, however, objected that including rules that adjust compositionality to fit extreme contextual sensitivity does not work. In his view, clear and stable contextual parameters cannot fix this as a type of sensitivity, for at least two reasons:

1. The meaning of a sentence in natural language depends on many factors and can vary widely.
2. There is no well-determined limit on the number of pragmatic factors the emission context can influence.

If this is the case, then it seems to be impossible to have a rule for weak compositionality. One would not always be able to predict which types of pragmatic phenomena will influence a sentence's content.

Briciu has raised such a concern from a different perspective: "If natural languages are compositionally weak and their expressions are radically context-sensitive, a single syntactic structure will contribute in more than one way to the interpretation of complex expressions". (2020, 217)

Consider the sentence:

John observed a man using his computer **[3].**

Here, we would usually say that there is a single syntactic structure with at least two interpretations: (i) John observed a man through his computer or (ii) a man was using John's computer. This shows that the algorithms which a weak compositionality rule obtains can hardly accommodate the degree of variability, owing both to extreme contextual sensitivity and to the sheer diversity of meanings that can be adjusted via pragmatic modulation. As Recanati suggests, weakening compositionality therefore does not actually have the benefits it hypothetically might have had. This makes it very complicated to retain compositionality in a theory that explains communication by accepting the breadth of linguistic meanings in natural language, so weakened compositionality fails to bring the benefits of compositionality to theories that study linguistic meaning in natural language. If weakening compositionality does not allow us to reconcile weakened and non-weakened compositionality, then there seems to be a problem with modifying our understanding of compositionality.

That said, one can argue that compositionality is valid and can be maintained, even if sentence meanings can be determined only subject to contextual dependence. Open compositionality makes it possible to explain how the meaning of a complex sentence is formed from its components (which belong to various domains besides the linguistic), so we can include a high degree of context sensitivity when examining a natural language sentence's meaning.

What is open compositionality?

When characterizing open-ended compositionality, one must take into account a thesis about the nature of language that García-Ramírez (2019) calls the "Lewisian Commitment"[10]: "Natural languages are, first and foremost, things that can be learned, developed and used by human beings given the limits and nature of their cognitive resources" (2019, p. 92). At first sight, such a commitment appears compatible with the formulation of compositionality explicated above. This is because open compositionality seeks to account for the characteristics that allow a language to be learned and its complex expressions to have meaning.

However, García-Ramírez has argued that the knowledge through which we learn a language stems from various epistemic dimensions[11]: not only those dependent on the language itself (as is the case with compositionality), but also cultural aspects, environmental stimuli, and emotional elements, among other things. What matters most when it comes to interpreting sentences is that, in open compositionality, we seek the most plausible interpretation of a complex expression in natural language. In other words, there are two requirements when interpreting a sentence's meaning: it must (i) be compatible with contextual variability and (ii) require little cognitive effort to interpret.

García-Ramírez states: "According to this view, compositional processes are only one among multiple different procedures required to account for the meaning of complex expressions" (2019, 16). There may be other ways, without compositional processes, to describe how complex sentences in natural languages have meaning. From a methodological point of view, open compositionality is a decision-making process in which we interpret meaning as economically as we can (i.e., based on speakers' informational constraints). Open compositionality also accepts contextual sensitivity because a sentence can convey different meanings depending on its context (as discussed above).

Given all this, compositionality (as formulated here) fails to fulfill the Lewisian Commitment. No single source determines anyone's knowledge of language. This definition of open-ended compositionality can be stated more rigorously as follows: Depending on contextual demands, the procedure of interpreting a sentence may sometimes involve heuristic strategies, syntactic algorithms, or both (either in parallel or as mutual backup strategies).

From this perspective, the context of a sentence[12] is what determines the meaning of a complex expression. Let us return to **[1]** in its three occurrences:

1. *In a park*. It would have a conventional meaning, namely that there is a lame duck and it probably needs help.

In this case, the complex meaning is determined by the conventional meanings of the words and by information about the world (the physical environment).

2. *In a business environment in the United States*. Here, the situation is that a manager is about to step down.

In this case, the complex meaning is determined by semantic knowledge of idiomatic expressions related to business practices in American English. An important circumstance also merits the use of "lame duck."

3. *In a cricket match*. A batter has been dismissed without making runs, and this has happened uninterestingly.

In this case, the complex meaning is determined by knowledge of cricketing jargon and by the identification of an event that merits the adjective "lame."

According to the notion of open-ended compositionality, these complex meanings are shaped by the context, background knowledge, and physical environment in which a given sentence is used. What helps us to understand the meanings of each formulation is that "lame duck" is a polysemous expression. Its meaning can come from different sources, so identifying them helps us to discern which of its possible meanings it receives from them. By letting us clarify the type of heuristic strategies we use to interpret statements, this example also illustrates a criterion that determines how we make decisions when interpreting a sentence in natural language. We can find the most economical interpretation of each occurrence of **[1]** because open compositionality appeals to a criterion that lets us discern the processes for disambiguating an expression in natural language.

Open compositionality admits that (most of the time) a complex sentence's linguistic meaning cannot be determined solely by the meanings and syntactic arrangement of its parts. It requires many other processes, which may be compatible with context sensitivity. Now, we are seeking a contextualist theory that considers such sensitivity while accounting for complex sentences' configurations of meaning, so we should appeal to some form of open compositionality. If radical contextualist theories are based on open compositionality, then they can overcome the problems classic compositionality faces, as I showed above. This allows me to formulate the following metatheoretical claim:

A radical contextualist theory can include compositionality if and only if that theory is based on a notion of sentence meaning that is configured according to open compositionality.

In this case sentence meaning is the most cognitively economical interpretation of a sentence in everyday discourse that is formulated at some specific moment in time and determined by some decision-making process.

Conclusions

I have presented a critical description of compositionality's formulation, some of its applications, some variations, and the most frequent objections. I have also discussed compositionality and radical contextualism. In so doing, I have provided a characterization of radical contextualism in general—one that may be compatible with various manifestations of radical contextualism. I employed a relevant notion of context to clarify this.

The need for compatibility between compositionality and radical contextualism involves the idea that a complex expression's meaning comes from its parts and syntactic arrangement—taking contextual sensitivity into account makes this difficult. Recanati's proposal to weaken compositionality by making it a function of sentence context fails because it attempts to cohere context-sensitive lexical items, which can hardly be treated formally. The main tension between classical compositionality and contextualism lies in contextual sensitivity. We should modify the notion of classical compositionality to endorse open compositionality when formulating radical contextualist theories.

Open compositionality implies that complex sentence meanings are determined by a series of processes—not necessarily linguistic ones—in which we choose the meaning that presents an interlocutor with the least cognitive effort. Such processes include the contextual variation to which sentence meanings are subject in any natural language, so compositionality can help a contextualist theory only if it is open compositionality. To characterize this, I have proposed a concomitant metatheoretical principle.

It will be useful to study this principle in future research, especially when one considers open compositionality's relevance to radical contextualist theories that are based on the need for contextual adjustment.

Notes

[1] Despite this attribution, finding a canonical definition of the compositionality principle in Frege's work is not easy. According to Janssen, "[t]he most well-known fragment of Frege which reminds of compositionality is the first sentence of *Gedankegefüge* [*Logical Investigations*] in the translation by Geach & Stoothoff [Geach & Stoothoff, 1977]: It is astonishing what language can do. With a few syllables, it can express an incalculable number of thoughts so that even a thought grasped by a terrestrial being for the very first time can be put into the form of words which will be understood by someone to whom the thought is entirely new. This would be impossible if we were not able to distinguish parts of the thoughts corresponding to the parts of a sentence so that the structure of the sentence serves as the image of the structure of the thoughts" (Janssen, 2001).

[2] Given this paper's metatheoretical nature, I shall limit myself to a very general analysis of compositionality. For a more detailed study, see Szabó (2012), who puts forward three compositionality theses in different disciplines: philosophy, linguistics, and psychology.

[3] Pagin (2012) mentions a second problem. He argues that learnability (as the argument conceives it) does not require compositionality, only that languages be computable. This topic, however, is orthogonal to my discussion in this paper.

[4] Peter Lasersohn (2012) discusses the reasons for this argument—an argument he does not find entirely convincing.

[5] For more on radical contextualist theories, see Carston (2012) and Travis (2008).

[6] David Lewis' notion of context in compositional semantics has also been highly influential. He states that a "context is a location – time, place, and possible world – where an utterance is said" (Lewis,1997; see Stalnaker, 2018 for more).

[7] I have chosen this notion because it is compatible with a general thesis underpinning radical contextualism and can include diverse theories.

[8] I have adjusted this rule to fit the terminology used in this paper without considering the definitions Recanati uses in his theory.

[9] Another way to characterize compositionality is as follows: maintain strict compositionality but separate it from the truth values of any interpretation that depends on how speakers judge the truth or falsity of an utterance. In this approach, we have compositional semantics that focuses on the syntactic and lexical mechanisms that formally represent meanings in a language. In other words, there is an abstraction of a particular language's meaning —an abstraction that excludes speakers' use of linguistic expressions.

[10] David Lewis (1975) established two perspectives on human language. The first perspective is formal— a language is an abstract entity consisting of a set of marks and sounds, independent of their use. From the

other perspective, human populations are guided by conventions of truthfulness and trust in using these marks and sounds. According to García- Ramírez, Lewis argued that human languages can be understood in purely formal terms. However, Lewis did not actually develop this commitment. Thus, one of the motivations behind open compositionality involves carefully configuring the Lewisian Commitment.

[11] García-Ramírez appeals to empirical evidence to support his argument for open compositionality. However, in this paper I assume that García-Ramírez's theoretical apparatus is plausible without carefully evaluating this evidence (see García-Ramírez, 2019, Ch. 3).

[12] Understood in the sense proposed by Pollock (2020).

References

Briciu, A. (2020). Compositionality in Truth-Conditional Pragmatics. In T. Ciecierski and P. Grabarczyk (Eds.), *The Architecture of Context and Context-Sensitivity* (pp. 205–225). Springer.

Carston, R. (2012). Word Meaning and Concept Expressed. *The Linguistic Review*, 29(4), 607–623.

Ciecierski, T. & Grabarczyk, P. (2020). An Introduction: The Architecture of Context and Context-Sensitivity. In T.G. Ciecierski and P. Grabarczyk (Eds.), *The Architecture of Context and Context-Sensitivity* (pp. 205–225). Springer.

Davidson, D. (2001). Theories of Meaning and Learnable Languages. In D. Davidson, *Inquiries into Truth and Interpretation* (pp. 1–15). Oxford University Press.

Dowty, D., Wall, R. & Peters, S. (1981). *Introduction to Montague Semantics*, no. 11 in Synthese Language Library, Reidel.

Elbourne, P. (2011). *Meaning: A Slim Guide to Semantics.* Oxford University Press.

García-Ramírez, E. (2019). *Open Compositionality: Toward a New Methodology of Language.* Lexington Books.

Geach, P. T. & Stoothoff, R. H. (1977). *Logical Investigations*. Gottlob Frege. Basil Blackwell.

Huang, M. (2017). A Plea for Radical Contextualism. *Synthese*, 194(3), 963–988.

Janssen, T. M. V. (2001). Frege, Contextuality and Compositionality. *Journal of Logic, Language and Information* 10(1), 115–136.

Jaszczolt, K. (2018). *Meaning in Linguistic Interaction: Semantics, Metasemantics, Philosophy of Language.* Oxford University Press.

Jaszczolt, K. (2023). *Semantics, Pragmatics, Philosophy: A Journey through Meaning*. Cambridge University Press.

Lasersohn, P. (2012). Contextualism and Compositionality. *Linguistics and Philosophy*, 35(2), 171–189.

Lepore, E. & Ludwig, K. (2009). *Donald Davidson's Truth-Theoretic Semantics.* Clarendon Press.

Lewis, D. (1975). Languages and Language. In K. Gunderson (Ed.), *Language, Mind, and Knowledge* (pp. 3–35). University of Minnesota Press.

Lewis D. (1997). Index, Context, and Content. In: *Papers in Philosophical Logic*. Cambridge Studies in Philosophy (pp. 21–44). Cambridge University Press.

Pagin, P. (2012). Communication and the Complexity of Semantics. In W. Hinzen, M. Werning, and E. Machery (Eds.), *The Oxford Handbook of Compositionality* (pp. 510–529). Oxford University Press.

Pollock, J. (2020). Occasion-Sensitivity and What Is Said. In T. Ciecierski and P. Grabarczyk (Eds.), *The Architecture of Context and Context-Sensitivity* (pp. 245–264). Springer.

Recanati, F. (2010). *Truth-Conditional Pragmatics*. Oxford University Press.

Stalnaker, R. (2018). David Lewis on Context. In D. Ball and B. Rabern (Eds.), *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics* (pp. 98–112). Oxford University Press.

Szabó, Z. G. (2012). The Case for Compositionality. In W. Hinzen, E. Machery, and M. Werning (Eds.), *The Oxford Handbook of Compositionality* (pp.64 –80). Oxford University Press.

Travis, C. (2008). *Occasion-Sensitivity: Selected Essays*. Oxford University Press.

Zimmerman, T. E. (2020). Fregean Compositionality. In D. Ball and B. Rabern (Eds.), *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics* (pp. 276–305). Oxford University Press.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 136-152

Artículo/Article

El lenguaje y "lo místico" en el primer wittgenstein

Language and "the mystical" in early wittgenstein

Javier Antonio Torres-Vindas

Universidad Estatal a Distancia (UNED), Costa Rica
jtorres@uned.ac.cr
https://orcid.org/0000-0002-0130-5979

DOI https://doi.org/10.48204/2805-1815.8477

INFORMACIÓN DEL ARTÍCULO

Recibido el: 14/7/2025

Aceptado el: 18/9/2025

Keywords:

Language, ethics, aesthetics, logical form, ineffable

Palabras clave:

Lenguaje, ética, estética, forma lógica, inefable

ABSTRACT/RESUMEN

Abstract:

This article examines the relationship between language and "the mystical" in Wittgenstein's Tractatus Logico-Philosophicus, emphasizing its opaque nature and its ethical-epistemic implications. First, it analyses the pictorial theory of language: propositions are logical figures (Bild) that share a logical form (Logische Form) with worldly facts, yet their limits prevent the expression of an "ultimate meaning." Second, it explores "the mystical" as the ineffable: the very existence of the world (6.44) and its perception as a bounded totality (6.45), which can only be shown (6.522), not stated. Third, it derives ethical-aesthetic consequences: the metaphysical subject, as the limit of the world (5.632), attains happiness through an aesthetic contemplation that accepts the world without intervention (will outside the world, 6.373). Ethics and aesthetics converge in a contemplative stance toward the mystical, whose logical inexpressibility demands silence (7). Thus, the Tractatus transcends positivism by pointing toward the transcendental through its own self-limitation.

Resumen:

Este artículo analiza la relación entre lenguaje y "lo místico" en el Tractatus Logico-Philosophicus de Wittgenstein, destacando su carácter opaco y sus implicaciones ético-epistémicas. Primero, examino la teoría pictórica del lenguaje: las proposiciones son figuras lógicas (Bild) que comparten forma lógica (Logische Form) con los hechos mundanos, pero su límite impide expresar un "sentido último". Segundo, exploro "lo místico" como lo inefable: la existencia misma del mundo (6.44) y su percepción como totalidad limitada (6.45), que solo se muestra (6.522), no se dice. Tercero, derivo consecuencias ético-estéticas: el sujeto metafísico, como límite del mundo (5.632), alcanza la felicidad mediante una contemplación estética que acepta el mundo sin intervenir (voluntad fuera del mundo, 6.373). La ética y estética

	convergen en una actitud contemplativa ante lo místico, cuya inexpresabilidad lógica exige silencio (7). Así, el Tractatus trasciende el positivismo al apuntar a lo trascendental mediante su propia autolimitación.

Introducción

Pocos pensadores, en tan pocas páginas pueden llegar a marcar época. Ludwig Wittgenstein (1889-1951), fue uno de ellos con su *Tractatus Logico-Philosophicus* (Wittgenstein, 1994). El presente artículo versa sobre el papel de “lo místico” y su relación con el lenguaje en dicha obra del autor, para ello haré tres acercamientos. Primero expondré su concepción de lenguaje, para pasar a realizar un acercamiento a su concepto de “lo místico”. Finalmente, con lo expuesto mostraré sus implicaciones ético-epistemológicas.

El *Tractatus* ha adquirido múltiples significados según sus diversos intérpretes. Su estructura se organiza mediante aforismos con un sistema de numeración jerárquico: cada sección principal lleva un dígito (del 1 al 7), las reflexiones sobre el primer aforismo se identifican como 1.1, 1.2, y los análisis posteriores de estos comentarios se codifican como 1.1.1, 1.1.2, 1.1.3, y así sucesivamente.

El *Tractatus* presenta sus aforismos como síntesis condensada de reflexiones que Wittgenstein desarrolló en sus cuadernos, material que fue articulando y depurando mediante múltiples reformulaciones (en ocasiones decenas de versiones). Aunque la obra destaca por su formato reducido (menos de 80 páginas), su lectura exige reiterados abordajes y una inversión temporal desproporcionada respecto a su extensión física, demandando horas de análisis meticuloso para desentrañar su densidad conceptual. Sus tesis por tanto son:

1. El mundo es todo lo que es el caso
2. Lo que es el caso, el hecho, es el darse efectivo del estado de cosas
3. La figura lógica de los hechos es el pensamiento
4. El pensamiento es la proposición con sentido
5. La proposición es una función veritativa de las proposiciones elementales (La proposición elemental es una función veritativa de sí misma)
6. La forma general de la función veritativa es [-p, -î N(-î)]. Esta es la forma general de la proposición.

7. De lo que no se puede hablar hay que callar.

Primer acercamiento: El lenguaje y el mundo

Para Wittgenstein, el lenguaje o su equivalente "el pensamiento" es finalmente, doctrina sobre la realidad, sobre el mundo (estados de cosas). Ya que la totalidad del espacio lógico abarca el mundo, lo describe, lo muestra asomando su sentido, pero el mismo es el límite opaco que no nos dice "sentido último". Entonces, los supuestos primordiales del *Tractatus* son:

- El isomorfismo del lenguaje y el mundo
- La reducción del lenguaje a su función descriptiva.

La capacidad de esta teoría del lenguaje (ideal) para establecer un isomorfismo entre el binomio lenguaje/mundo radica en la presencia de un elemento compartido: su forma. La forma lógica del lenguaje (*Logische Form*) puede revelarnos la forma lógica de la realidad, ya que ambas son en esencia la misma.

Ahora, una teoría del lenguaje que refleja el mundo o bien una teoría del mundo que se refleja en el lenguaje implica la construcción de límites. El único modo de trazar límites al pensamiento y con ellos al mundo es a través de su expresión: el lenguaje. (Wittgenstein, 1994, 5.6)

La teoría del lenguaje se funda en la necesidad de llegar a desplegar proposiciones elementales, cuyo análisis final degenera en los nombres que tendían correspondencia con los objetos "atómicos" (indivisibles) del mundo.

Por tanto, los objetos en el *Tractatus* son una exigencia lógica que responde a la necesidad de llegar a un término final en el proceso de análisis. Los objetos, tienen como característica esencial la posibilidad de ser parte constitutiva de los estados de cosas (Wittgenstein, 1994, 2.02).

Los objetos son simples: ingenerables, indestructibles e independientes (causalmente) Esto, en tanto, es una responde al doble nivel ontológico-epistémico del Wittgenstein del *Tractatus*: la representación del estado de cosas mediante figuras requiere la existencia de objetos (metafísicamente) simples. El objeto contiene todas las posibilidades de entrar en combinación con otros objetos (Wittgenstein, 1994, 2.0121), determinando así el estado de cosas de los que puede formar parte (Wittgenstein, 1994, 2.0123).

Es decir, Wittgenstein introduce una distinción del estado de cosas, “donde lo fijo e invariable son los objetos que, en efecto, existen con independencia de los hechos en que toman parte. Lo variable es la configuración de los objetos, es decir, los hechos” (Wittgenstein, 1994, 2.026-2.0271). Acusando más finamente, hemos de rescatar: lo que las proposiciones describen son el estado de cosas (Wittgenstein, 1994, 4.023) y por ello, los objetos no se pueden expresar, sólo se les puede nombrar (Wittgenstein,1994, 3.221).

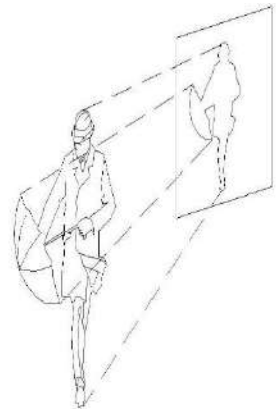
Por lo tanto, al clarificar la naturaleza del lenguaje, se reflejará de manera automática la naturaleza del pensamiento y de la realidad. El estudio del lenguaje se constituye así en condición necesaria/suficiente del conocimiento del mundo: (Sí)... Los hechos en el espacio lógico son el mundo (Wittgenstein, 1994, 1.13) (entonces)... Lo que no podemos pensar no lo podemos pensar; así pues, tampoco podemos decir lo que no podemos pensar (Wittgenstein,1994, 5.61) (por tanto) ... la forma lógica de la proposición es: las cosas se comportan de tal y tal modo (Wittgenstein,1994, 4.5) (en consecuencia) ... lo inexpresable ciertamente, existe. Se muestra, en lo místico (Wittgenstein,1994, 6.521).

La forma clave de articulación para comprender dicha teoría del lenguaje se logra a través de su Teoría Pictórica de la Proposición. Dicha tesis refiere el lenguaje es un retrato lógico (pintura) del mundo. Por tanto, el lenguaje asienta como descriptivo de los hechos: “el mundo es todo lo que es el caso”. Ver figura 1.

Así, la realidad del mundo se define como la suma total de los hechos existentes, mientras que el lenguaje se configura como el conjunto completo de proposiciones, donde cada una de ellas representa una posible configuración fáctica del mundo. Podemos afirmar:

- La proposición (*Satz*) es una figura (*Bild*) de la realidad
- El lenguaje puede figurar la realidad (mundo) porque ambos comparten la misma forma lógica (*Logische Form*).
- El lenguaje y el mundo (estado de cosas) cumplen una relación figurativa (*Abbildende Beziehung*) en tanto existe correspondencia (descriptiva) entre lo figurado y la figura.
- A la proposición pertenece todo cuanto pertenece a la proyección; pero no lo proyectado (Wittgenstein,1994, 3.13)

Figura 1
Diagrama esquemático de la “Teoría pictórica de la proposición”



Fuente: Elaboración propia

Ahora, el lenguaje concatena o articula las proposiciones elementales para formar entramados proposicionales complejos. Las proposiciones elementales poseen carácter afirmativo constitutivo, pues la negación se configura como operación lógica derivada que genera funciones veritativas mediante procesos de combinación. En su estado primario, estas proposiciones manifiestan positividad ontológica, mientras que la negación emerge como construcción lógica secundaria donde la falsedad se obtiene mediante operaciones de inversión sobre los valores de verdad originales.

Si nos hacemos figuras de los hechos (Wittgenstein,1994, 2.1) es decir, proposiciones; el lenguaje se constituye como la totalidad de las proposiciones (Wittgenstein,1994, 4.001), lo que supone que el lenguaje en su conjunto describe no sólo el mundo real, sino, el mundo posible, formado por el conjunto de los hechos que pueden acaecer (Wittgenstein,1994, 5.6)

Es decir, todo el espacio lógico y por tanto la totalidad del lenguaje es una tautología que se desdobla así misma, lo que se traduce en la imposibilidad de salir del lenguaje. En otras palabras; la (*Form der Abbildung*) figuración del lenguaje (Wittgenstein,1994, 2.17) no puede figurar su propia forma de figuración, la ostenta (Wittgenstein,1994, 2.171), la porta, la muestra (*Zeigen*). Como señala Zemach (1966/1997):

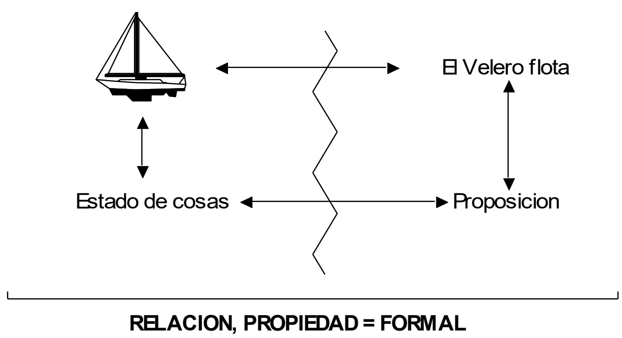
> ... una de las ideas centrales del *Tractatus* es la diferencia esencial entre contener (*enthalten*) y tener (*haben*). Una proposición, por ejemplo, tiene un sentido, lo presenta (*stelltdar*). Pero no lo contiene (2.203, 3.13) Un sentido (*un Sachlage*) es expresado

> por una proposición, pero no está el mismo incorporado a la proposición... Lo figurado, el sentido, está fuera de la proposición que lo expresa. (p. 18)

La proposición muestra su sentido, muestra cómo se comportan las cosas si es verdadera. Y dice que se comportan así (Wittgenstein,1994, 4.022) La figura 2 representa su objeto desde fuera (su punto de vista es su forma de representación); por ello representa su objeto correcta o falsamente (Wittgenstein,1994, 2.173) (pues) el método de proyección es el pensar el sentido de la proposición (Wittgenstein,1994, 3.11). Ver figura 2.

Figura 2

Diagrama esquemático de la representación de mundo en el lenguaje



Fuente: Elaboración propia

Ergo, la representación de estados de cosas mediante figuras exige la existencia de elementos ontológicamente simples (no meramente lógicos), donde la relación no se reduce a una equivalencia formal entre sintaxis (categoría de nombres) y semántica (categoría de objetos). La condición fundamental radica en la naturaleza metafísica de los objetos simples como requisito previo para la posibilidad misma de la representación figurativa, trascendiendo así una mera correspondencia categorial entre lenguaje y realidad.

Segundo acercamiento: "Lo mítico"

El problema de "lo místico" en el *Tractatus*, Wittgenstein (1994) lo plantea en los siguientes términos:

- 6.44: No es lo místico como sea el mundo, sino que sea el mundo.
- 6.45: La visión del mundo *sub specie aeterni* es su contemplación como un todo -limitado-. Sentir el mundo como un todo limitado es lo místico.

- 6.5: Para una respuesta que no se puede expresar, la pregunta tampoco puede expresarse. No hay enigma. Si se puede plantear una cuestión, también se puede responder.
- 6.522: Hay, ciertamente, lo inexpresable, lo que se muestra a si mismo; esto es lo místico.
- 7: De lo que no se puede hablar, mejor es callarse.

El misticismo en el *Tractatus* trasciende lo discursivo al abordar el sentido del mundo, articulándose en dos dimensiones clave:

1. Se configura como una indagación que concibe el mundo como totalidad finita, donde la limitación ontológica impulsa la pregunta por el significado último o la ética como búsqueda del sentido
2. La solución al enigma del sentido no reside en proposiciones lógicas, sino en la contemplación estética del mundo como *todo unificado*, donde lo místico emerge como experiencia límite que desborda los marcos del lenguaje o la estética como clave hermenéutica.

Lo místico se define como vivencia no proposicional: las aproximaciones místicas, aunque revestidas de aparente profundidad, renuncian a la explicación para señalar lo radicalmente inexpresable. Su función no es decir sino mostrar la paradoja de un sentido que solo puede experimentarse en el silencio de lo indecible.

El *Tractatus* postula una identidad estructural absoluta entre lenguaje y realidad, donde la lógica opera como esqueleto transcendental que determina tanto lo decible como lo existente. Este isomorfismo no se reduce a una mera correspondencia categorial (nombres-objetos, proposiciones-hechos), sino que implica que la forma lógica (*Logische Form*) es el sustrato común que posibilita la representación.

Las proposiciones elementales, al carecer de operadores lógicos, revelan esta simetría primigenia: su configuración afirma directamente estados de cosas posibles, mientras que la negación emerge como función veritativa derivada. Así, el lenguaje no solo describe el mundo, sino que comparte su arquitectura metafísica, circunscribiendo lo pensable dentro de los límites de lo factual.

Wittgenstein construye su sistema lógico con un propósito autoaniquilante: demostrar que lo esencial (ética, estética, sentido del mundo) yace más allá de lo representable. Al delimitar rigurosamente el lenguaje a lo factual, expulsa lo místico al territorio de lo inefable, donde solo puede *mostrarse* mediante experiencias límite, nunca *decirse*.

Esta estrategia genera una tensión irresoluble: el propio *Tractatus* emplea proposiciones para señalar su insuficiencia, incurriendo en lo que él mismo denuncia como sinsentido (*Unsinning*). La solución radica en comprender el texto como escalera que debe ser desechada tras su uso: su valor no está en sus afirmaciones, sino en la indicación silenciosa de lo trascendente.

La dimensión mística culmina en una ética negativa: al reconocer que los valores no son hechos mundanos, se transforman en condiciones de posibilidad que estructuran la mirada ética sin poder articularse. Esta ética no prescribe normas, sino que se manifiesta en la actitud ante el mundo como totalidad limitada, donde el sujeto ético experimenta el mundo *sub specie aeterni* (bajo la perspectiva de lo eterno).

El silencio resultante no es vacío, sino pleno de significado: constituye el espacio donde lo místico se actualiza como vivencia, no como discurso, estableciendo así una frontera infranqueable entre lo dicho y lo vivido.

El *Tractatus* se presenta como un estudio "lógico" que busca desarrollar un lenguaje claro y preciso para clasificar pensamientos e ideas, evitando así los problemas filosóficos que carecen de sentido. Por lo tanto, el objetivo de la filosofía es la clasificación lógica de los pensamientos. El resultado de esta labor no consiste en un conjunto de “proposiciones filosóficas”, sino en la clarificación de dichas proposiciones.

La filosofía debe aclarar y delimitar cuidadosamente los pensamientos, ya que, de lo contrario, estos tienden a ser confusos e imprecisos (Wittgenstein,1994, 4.112). Esto solo se puede lograr mediante la depuración del lenguaje lógico, ya que la filosofía "debe establecer lo que es pensable y, por ende, lo que no lo es" (Wittgenstein,1994, 4.114). A través del análisis lógico, podemos descubrir lo que subyace a la realidad, es decir, la esencia del mundo.

Wittgenstein se interesa en identificar los elementos fundamentales del lenguaje. Según él, el lenguaje actúa como una imagen de la realidad, una representación que "manifiesta" el mundo. Este mundo se basa en los "hechos", que son las unidades mínimas de la existencia: "Lo que sucede, el hecho, son las unidades más básicas del mundo" (Wittgenstein,1994, 2).

Estos hechos son las unidades mínimas del mundo y tienen una entidad lógica, no física. En el mundo, está claro, hay objetos, pero éstos por sí mismos no son nada, ya que en el mundo lo que hay son hechos, conjuntos de sucesos; por ejemplo, decimos "esto es blanco" o "está lloviendo".

De la misma manera que en el mundo lo que es significativo es el "hecho" y no el "objeto", en el lenguaje lo que es significativo son las "proposiciones" y no sus elementos, por esto "Sólo la proposición tiene sentido; sólo en el contexto de la proposición tiene significado un nombre" (Wittgenstein,1994, 3.3). Esto será entendido como la "teoría figurativa del sentido".

Ergo, el *Tractatus* presenta una contradicción tanto teórica como práctica en relación con la proposición 7, que se considera su proposición fundamental. En su primera parte, esta proposición sugiere que existen aspectos de la realidad sobre los cuales no se puede hablar. Sin embargo, en su segunda parte, se insta a no abordar estos temas. Esta dualidad plantea un dilema sobre la capacidad del lenguaje para abordar lo inefable y desafía la coherencia de su propio marco filosófico.

Es un hecho que genera confusión. ¿Qué significa que una mente tan lógica no haya seguido su propia regla y, al hacerlo, haya invalidado ya sea esa regla o todo lo que contradice dicha regla?

El dilema no se resuelve con la afirmación al final del tratado (Wittgenstein,1994, 6.54) de que sus proposiciones son como una escalera que debemos desechar una vez que hemos logrado una comprensión adecuada del mundo, ya que no es tan evidente que el resultado se pueda separar del medio que permite alcanzarlo.

¿Es ese medio simplemente una herramienta que puede ser reemplazada por cualquier otra, de tal manera que cualquiera podría llevar al mismo resultado? Si fuera así, ¿por qué, entonces, la escalera que el lector del *Tractatus* debe utilizar y que debe desechar después de usar no es cualquier otra escalera, sino precisamente la que Wittgenstein empleó, su propia escalera? Este dilema no tiene solución dentro del *Tractatus*.

Sin embargo, el autor, se ve obligado a salir, fuera de los límites que él mismo ha impuesto; pero esta transgresión de sus propias leyes puede ser excusada en virtud de su necesidad metódica. En este sentido, Russell en su introducción al *Tractatus* afirma:

> Más interesante que estas cuestiones de detalle comparativo es la actitud de Wittgenstein respecto a lo místico. Su actitud hacia ello nace de modo natural de su doctrina de la lógica pura, según la cual, la proposición lógica es una figura (verdadera o falsa) de hecho, y tiene en común con el hecho cierta estructura. (1994, p. 196)

Es precisamente esta estructura compartida entre lenguaje y realidad lo que permite al primero representar hechos. Sin embargo, dicha estructura no puede verbalizarse, pues constituye el esqueleto lógico tanto del lenguaje como de los hechos que describe. Por ello, todo lo relacionado con la capacidad expresiva del lenguaje resulta inefable en un sentido riguroso: incluye la totalidad de la lógica y la filosofía misma.

Wittgenstein postula que nuestra comprensión trasciende lo verbalizable. Aunque visualizamos los confines del lenguaje, lo que yace más allá –las condiciones de posibilidad del sentido– no puede formularse en proposiciones. Los objetos solo admiten ser nombrados (Wittgenstein,1994, 3.221), nunca definidos. Las proposiciones describen *cómo* son las cosas, no *qué* son. La exigencia de signos elementales (Wittgenstein,1994, 3.23) garantiza la determinación del sentido: sin ellos, el lenguaje carecería de anclaje en lo concreto.

La existencia de objetos como sustrato ontológico (Wittgenstein,1994, 2.026-2.027) es requisito fundamental. Sin esta "forma inalterable", el sentido se desvanecería en la indeterminación, imposibilitando tanto la construcción de imágenes del mundo como su descripción. Paradójicamente, esta condición transcendental de todo lenguaje descriptivo permanece ella misma fuera del ámbito de lo descriptible. Esta reformulación mantiene los núcleos conceptuales originales mientras:

1. Jerarquiza las ideas mediante subtítulos,
2. Simplifica la sintaxis sin perder rigor técnico,
3. Introduce términos equivalentes ("sustrato ontológico" por "forma inalterable"),
4. Explicita las implicaciones filosóficas de las tesis wittgensteinianas.

Lo místico, lo inefable no es algo “fuera” del mundo (extramundo) sino más bien sobre o supramundano. Está por encima de la lógica. Lo inefable es lo que está más arriba del mundo y del yo (fenoménico, psicológico, metafísico) lo envuelve, lo configura. Así, se indica “El sentido del mundo tiene que residir fuera de él. En el mundo todo es como es y todo sucede como sucede; en él no hay valor alguno, y lo hubiera carecería de valor” (Wittgenstein,1994, 6.41). Ergo, el concepto de “lo místico” se refiere (en el *Tractatus*) al sentido del mundo en su totalidad (limitada) no cómo el mundo sea, sino a su existencia.

La reflexión de Wittgenstein sobre el mundo se centra en la distinción entre dos órdenes: el reino de los hechos, que se refiere a lo decible y contingentemente verdadero, y el ámbito del sentido, que abarca lo mostrable y lo trascendental. La paradoja fundamental radica en que el significado del mundo no puede ser un hecho en sí mismo; más bien, la "razón de ser" última actúa como condición de posibilidad para todos los hechos, manifestándose en ellos sin ser un hecho. Wittgenstein propone una solución contemplativa que se articula en tres movimientos filosóficos: el ético, que reconoce los límites del lenguaje; el estético, que percibe el mundo como una totalidad; y el místico, que vive lo inexpresable como el horizonte de todo significado.

Además, el conocimiento factual, representado por la ciencia, resulta insuficiente para abordar lo esencial, transformando la filosofía en una actividad terapéutica que apunta hacia lo inefable. La ética se convierte en una actitud ante la existencia, desvinculándose de la moral convencional. El *Tractatus* refleja esta tensión al utilizar proposiciones lógicas para mostrar lo que no puede expresarse, actuando como un espejo de la estructura del mundo. Su autodisolución final (1994, 6.54) enfatiza la relación entre lenguaje y sentido, sugiriendo que la comprensión auténtica del mundo es existencial, implicando un reajuste de nuestra mirada que transforma nuestra relación con la realidad.

Tercer acercamiento: Implicaciones ético-epistemológicas

Como hemos visto, las proposiciones lógicas, aunque no están tan directamente relacionadas con el mundo como las de la ciencia natural que reproducen la realidad (y tienen sentido), siempre son verdaderas y muestran la verdad, por esto hemos de fundamentarnos en ellas (vemos aquí el carácter claramente positivista de este argumento).

La "forma lógica" es la constitutiva del lenguaje y no puede ser reproducida con "sentido" porque es la que permite el propio lenguaje ("sólo se puede mostrar en la proposición misma"). Por esto advierte Wittgenstein (1994): no se puede ir más allá de la realidad, cuando entiendo que es lo que es más elevado.

Así, no tiene ningún sentido preguntarse si hay algo más allá del mundo o buscar sentidos más allá de él. Wittgenstein nos está diciendo que no puede manifestarse nada más elevado (proposiciones estéticas, éticas, metafísicas...) de lo que realmente hay aquí, en el mundo, que puede quedar expresado por el lenguaje en todo caso.

Si hay alguna cosa más que tiene sentido, así, es aquí, no más allá; el sentido "místico", por tanto, no está en mundos más elevados. Pero esto es difícil de aceptar "-no tendría la sensación de que le estábamos enseñando filosofía-" (Wittgenstein,1994, 6.53), aunque es lo que realmente es "correcto". Esto es lo que habría, pues, de decir la filosofía, "no decir más que aquello que se puede decir, o sea, proposiciones de la ciencia natural..." (Wittgenstein,1994, 6.53)

La religión, la estética, la ética, toda actitud dirigida hacia los valores absolutos, toda pregunta por el “sentido” del mundo y de la vida, choca con los límites del lenguaje y se basa en la captación del mundo como portento inexplicable (Cf. Villidoro, 1975, p. 13).

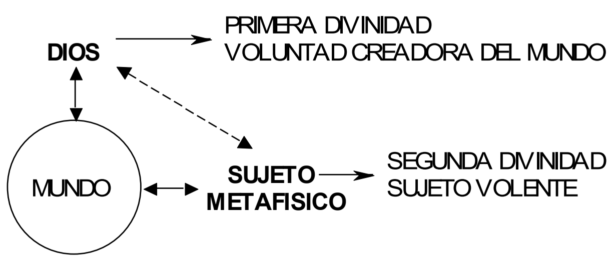
El problema primario que enfrenta lo ético-estético en Wittgenstein se expresa en: “No es lo místico como sea el mundo, sino que sea el mundo” (1994, 6.44). La existencia se explaya en el *Tractatus* en al menos tres categorías de existir:

a. La primera se refiere a la existencia del mundo como tal: No cómo sea el mundo es lo místico, sino que sea (1994, 6.44)
b. La existencia del mundo independiente de mi voluntad (1994, 6.373) donde mi yo es un límite del mundo (1994, 5.632 y 5.633) y, por tanto, soy microcosmos (1994, 5.63)
c. La existencia como misterio inefable, infinito e ilimitado; donde la muerte no es un acontecimiento de la vida (1994, 6.4311) y mi actitud hacia la existencia puede ser feliz/infeliz (1994, 6.43); es decir, cambia el límite del mundo y su voluntad hacia la existencia.

¿Quién enfrenta este problema? El sujeto metafísico[1] (Ego Trascendental), o sea, el sujeto último de todas las actividades y operaciones del individuo humano. Aquí no se puede hablar de ninguna experiencia directa interna o externa o de síntesis politética, sino de un saber indirecto, connotativo, o sea de un horizonte que nunca puede dejar de serlo, de un fondo que nunca aparece como figura. “El sujeto metafísico no pertenece al mundo, sino que es un límite del mundo” (Wittgenstein,1994, 5.632)

Ahora, este sujeto metafísico es su mundo (Wittgenstein,1994, 5.63), podemos volvernos por la reflexión sobre nuestro mundo interno para captar los actos del yo empírico, pero a la vez, en esta percepción apercibimos indirectamente el Yo trascendental que es el sujeto metafísico que reflexiona sobre esos actos y a la vez permanece como un horizonte oscuro, como fondo de todas las percepciones mundanas. Ver figura 3.

Figura 3
Diagrama esquemático de la relación sujeto-mundo



Fuente: Elaboración propia

Siguiendo a Zemach[2]: “El sujeto pensante da al mundo una forma. El sujeto volente le da un sentido – significación – ambas divinidades son trascendentales, esto es, constitutivas del mundo”. (1966, p. 18) ¿Cuál es la actitud del sujeto metafísico frente al mundo? Para Wittgenstein, es la contemplación del mundo. Es su percepción estética: El milagro estético es la existencia del mundo... Que exista lo que existe... ¿Es la esencia del modo de contemplación artístico contemplar el mundo con ojo feliz? Sería es la vida, alegre es el arte (Wittgenstein, 1982, p.145). En esta misma línea afirma Zemach (1966):

> El objeto del arte es así autosuficiente porque es una expresión del estatus último de ser un hecho. Es bello porque es una expresión de la factualidad, es decir, de la voluntad de Dios... la ética y la estética son lo mismo. Ellas son la expresión humana de asombro cuando él, el sujeto volente, encuentra lo místico: la

> existencia del mundo. “Estéticamente, que el mundo exista. Que lo que existe exista” (20.10.16) La ética y la estética son la manera mediante la cual la segunda divinidad se pone en armonía con la primera divinidad, lo místico. Ellas son la infinita aceptación de lo que es tal y como es. (pp. 20-21)

La apropiación del sentido del mundo consiste en hacerlo propio mediante una armonía existencial (Wittgenstein, 1994, 6.423-6.4311), donde la felicidad emerge como consecuencia de la virtud ética. Este estado se alcanza mediante una actitud contemplativa: el individuo, al renunciar a la voluntad de modificar un mundo irremediablemente marcado por el sufrimiento, asume una posición de observación pasiva. En esta perspectiva, la felicidad no radica en la acción transformadora, sino en la comprensión del significado último de la existencia, incluyendo el dolor, y en el reconocimiento del propio lugar como límite constitutivo del mundo. El sujeto ético, al trascender el deseo, accede a una visión totalizadora donde su posición marginal (como frontera del mundo) se revela como condición necesaria para captar el sentido trascendental de lo real.

El hombre feliz es inmortal. Desde el momento en que el sujeto se instala en los límites del mundo y se independiza de él; los hechos dejan de afectarle y con ellos también el tiempo. (Wittgenstein, 1994, 6.4311) Se supera la muerte con una eternidad absoluta (1994, 6.4312) y entrando en esta atemporalidad se está en actitud MISTICA = CONTEMPLATIVA:

> Lo místico en Wittgenstein se manifiesta como una dualidad esencial: por un lado, implica una percepción particular del mundo (captación de lo inefable), y por otro, una postura ética activa frente a él. Esta combinación transforma la experiencia metafísica en religiosa: mientras la dimensión contemplativa revela lo místico como horizonte de sentido, la actitud ética —de asombro y compromiso existencial— le otorga su carácter trascendente. Así, lo religioso no reside en dogmas sino en esta síntesis entre ver el mundo *sub specie aeternitatis* y actuar conforme a ese entendimiento, donde la ética opera como puente entre lo místico percibido y su expresión vital. (Cabrera, 1989, p.110).

Esta voluntad, al situarse fuera del mundo y carecer de capacidad para alterarlo (Wittgenstein, 1994, 6.373), no produce consecuencias factuales, sino que se inscribe en el propio límite del mundo, es decir, en la voluntad misma. La felicidad se vincula esencialmente a la perspectiva ética: aquellos que logran armonizar su existencia con el sentido del mundo —los "buenos"— alcanzan una visión feliz de la realidad (1994, 6.43). Esta correspondencia ético-existencial implica apropiarse del significado trascendental del mundo mediante una actitud contemplativa, donde la aceptación del límite (la voluntad como frontera) se convierte en la condición para superar la contingencia de los hechos y acceder a una comprensión plena de lo real.

Ética, estética, religión y felicidad no son en último término sino nombres para la misma realidad: la visión del mundo *sub specie aeterni.* Siendo la ética, esencialmente contemplación, la felicidad tiene un carácter semejante: estado pasivo de quien ha renunciado a querer y se limita a ver (Wittgenstein, 1982, 6.7.16) El mundo del feliz es, el de aquel que ha sometido todo deseo por lo contingente y ha alcanzado la abnegación.

Al instalarse en los límites del mundo, se independiza de él, los hechos dejan de afectarle, se hace inmortal (Wittgenstein, 1994, 6.4311 y 6.4312) Ahora lo místico es inexpresable (Wittgenstein, 1994, 6.5 y ss.) y la solución radica en que el lenguaje nos muestre su dimensión ostensiva (la ética, estética y la lógica son trascendentales). Lo místico es lo más importante (Wittgenstein, 1994, 7) La lógica es la herramienta formal para comprender el cuerpo que se esconde tras el ropaje del lenguaje, es la herramienta que nos libra de la voluntad que se afrenta al mundo, que nos libera del hombre infeliz y por ser límite es inefable[3].

Conclusiones

La imposibilidad de expresar lo místico es una imposibilidad lógica. La única salida es el silencio (Wittgenstein,1994, 6.5 y ss.): la solución al problema del mundo está en no intentar plantearlo y dejar que el lenguaje (sin su vestimenta, en sus formas puras) nos lo muestre en su dimensión ostensiva.

Lo místico es inefable no por absurdo, sino por importante, por infranqueable. La ética/estética es un modo de vida, no una teoría, y no puede ser mostrada con palabras, sólo mostrada con la conducta contemplativa del hombre. Es un cerrar y abrir los ojos ante lo místico de manera estética/ética, es como el instante de vigilia diurna es un mirar sin ver lo Otro. Es la experiencia del salto al abismo: lanzar gritos que repercuten en la profundidad del abismo (Nietzsche).

La existencia del mundo, con su surgimiento contingente y carente de fundamentación última, se revela como un fenómeno milagroso que desafía toda explicación causal. Wittgenstein señala que este *asombro metafísico* ante lo existente —lejos de ser una reacción ingenua— constituye una experiencia primordial que la racionalidad científica no logra disipar.

El verdadero estupor filosófico requiere un despertar epistemológico, pues el conocimiento convencional (atado a la red de hechos) opera como narcótico ante lo extraordinario de la mera existencia. Lenguaje y pensamiento, limitados a describir estados de cosas mediante proposiciones factuales, resultan radicalmente insuficientes para abordar el *milagro ontológico*: aquel puro "que haya mundo" en lugar de nada, misterio que se sustrae a toda lógica objetiva y se manifiesta como horizonte último de lo decible. (Cabrera, 1989, p. 111)

El proceso de Wittgenstein es claro (lógica/místico) despojar progresivamente al espíritu de todo lo que le es ajeno, de todo lo contingente: éxtasis y silogismo. ¿Detenerse ante la puerta en silencio y no traspasarla?

Notas

[1] El sujeto metafísico, que representa la totalidad del lenguaje, no puede ser parte del mundo, ya que este carece de elementos necesarios, salvo proposiciones lógicas vacías. Se concibe como una presuposición de voluntad frente al mundo, mientras que el "yo" se identifica con la suma de pensamientos. El yo del solipsismo, que abarca toda la realidad, no puede ser el sujeto empírico, que es solo un hecho del mundo. En el *Tractatus*, el sujeto metafísico es un límite del mundo y una necesidad lógica, ya que todos los hechos son contingentes y las proposiciones del lenguaje tienen condiciones de verdad. Así, el sujeto metafísico no puede formar parte de nuestra experiencia, que se limita a los hechos, sugiriendo que lo que existe podría ser de otra manera.

[2] *Filosofía de lo místico de Wittgenstein*, traducción libre e inédita de Mario Salas, Universidad de Costa Rica, 1997.

[3] La posición de Wittgenstein revela una paradoja existencial: en un mundo regido por la lógica de los objetos —donde todo enunciado significativo debe referirse a hechos comprobables—, lo místico (aquello que trasciende lo factual) queda condenado al silencio. Este mutismo no es accidental sino constitutivo: el filósofo, al vislumbrar realidades no objetivas en su conciencia (lo ético, lo estético, lo religioso), choca con los límites infranqueables del lenguaje. La tragedia surge cuando esta conciencia reconoce su radical inconmensurabilidad con el mundo: al igual que el héroe trágico clásico, el filósofo habita un abismo de incomunicación donde su comprensión del sentido último (captado *sub specie aeternitatis*) resulta intraducible al lenguaje de los hechos. Así, la existencia filosófica se vuelve un ejercicio de tensión permanente entre la certeza interior de lo inefable y la imposibilidad de compartirlo, generando una soledad ontológica donde toda pretensión de comunidad epistémica o compromiso existencial se revela como ilusión.

Referencias bibliográficas

Cabrera, I. (1989). Wittgenstein: la tentación de lo místico. *Diánoia,* 35(35), 107–116. https://doi.org/10.22201/iifs.18704913e.1989.35.636

Villoro, L. (1975). Lo indecible en el Tractatus. *Crítica,* 7(19), 19-31.

Wittgenstein, L. (1982). *Diario filosófico*. Editorial Ariel.

Wittgenstein, L. (1994). *Tractatus Logico-Philosophicus*. Editorial Alianza.

Zemach, E. (1966). Wittgenstein's Philosophy of the Mystical [La filosofía de lo mistico de Wittgenstein]. En I. M. Copi & R. W. Beard (Eds.), *Essays on Wittgenstein's Tractatus* (pp. 359-375). Routledge & Kegan Paul.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 170-181

Ensayo/Essay

Pensamiento crítico: una visión desde la filosofía

Critical thinkng: a philosophical view

José Antonio Mathurín

Universidad de Panamá, Panamá
josemathurin@gmail.com
https://orcid.org/0009-0007-4577-0446

DOI https://doi.org/10.48204/2805-1815.8480

INFORMACIÓN DEL ARTÍCULO

Recibido el: 24/2/2025

Aceptado el: 21/8/2025

Keywords:

Argument, citizenship, critical thinking, democracy, reason

Palabras clave:

Argumento, ciudadanía, pensamiento crítico, democracia, razón

ABSTRACT/RESUMEN

Abstract

The present article is a reflexive exercise, as well as a proposal, around nature, meaning, aims, value, and perspectives of critical thinking; the purely conceptual dimension of expression is distinguished its semantic scope, as well as that of practical normative nature, and its applicability. This reflection appears structured in five thematic axes articulated around a basic premise, namely: critical thinking is essentially philosophical thought. These thematic axes are expressed as follows: (i) Critical thinking: historical / cultural referential contexts. (ii) What is critical thinking: conceptual framework and conceptual precisions. (iii) The development of critical thinking: self-critical attitude and capabilities logical/argumentative, (iv) The levels of applicability of critical thinking. (v) Critical thinking, democracy, citizenship, and political order. Each of these components can be considered as an autonomous unit and independent, without losing its organic articulation with the rest.

Resumen

El presente ensayo es un ejercicio reflexivo, a la vez, una propuesta en torno a la naturaleza, sentido, fines, valor y perspectivas del pensamiento crítico. Se distingue la dimensión puramente conceptual de la expresión, es decir, su ámbito semántico, así como la de carácter práctico normativo y de aplicabilidad. Esta reflexión aparece estructurada en cinco ejes temáticos relacionados a una premisa básica: el pensamiento crítico es, en esencia, pensamiento filosófico. Estos ejes se expresan así: (i) El pensamiento crítico: contextos referenciales históricos/culturales; (ii) qué es (y no es) el pensamiento crítico: marco y precisiones conceptuales; (iii) el desarrollo del

	pensamiento crítico: la actitud autocrítica y las capacidades lógico/argumentativas; (iv) los niveles de aplicabilidad del pensamiento crítico; y (v) pensamiento crítico: La democracia, ciudadanía y el orden político. Cada componente puede considerarse autónoma e independiente, sin perder su articulación orgánica con el resto.

Introducción

El intento de ensayar una reflexión en torno al pensamiento crítico desde la una perspectiva filosófica no debe resultar contradictorio, toda vez que se parte de la premisa de que el pensamiento crítico ha de entenderse, fundamentalmente, como pensamiento filosófico. Ergo la filosofía es en esencia un enfoque crítico acerca de la realidad fenoménica; la filosofía, por antonomasia, es una labor crítica[1]: de ello da fe la propia historicidad de la filosofía.

Si admitimos esta premisa, consecuentemente reconocemos que el pensamiento crítico tiene por lo menos 2500 años de antigüedad[2]. Aparece en la cultura occidental[3] como producto de un largo, lento y tortuoso proceso resultante de la contraposición entre la visión racional del mundo y la permanente pugna con los poderes institucionales y tradicionales alimentados por el dogmatismo y el fundamentalismo.

Un ejemplo clásico de esa pugna lo encontramos en el sorprendente caso de los griegos, a quienes debemos los orígenes de la crítica, momento clave en la transición del sentido común a la ciencia, y de la necesaria distinción entre un saber carente de fundamento (acrítico) y el saber con fundamento racional (critico).

Este abordaje de la realidad y el mundo, que llamamos “pensamiento crítico”, se presenta reiteradamente a lo largo de la historia de la humanidad, pero siempre en permanente confrontación con los poderes instituidos (políticos, ideológicos, mercantiles, religiosos, etc.).

Qué es el pensamiento crítico

“Pensamiento crítico” resulta en extremo multívoco y polivalente, y –en algunos casos– vago y hasta arbitrario. Esto hace imprescindible un análisis conceptual para esclarecer el significado de este constructo.

En primer lugar, se debe revisar y distinguir dos tipos de pensamiento crítico, el que alude a una determinada forma o estilo de abordaje del mundo, la naturaleza y la sociedad; y el referido, por una parte, a una especial actitud o predisposición subjetiva y, por otra, al ámbito de aplicabilidad de operaciones y desarrollo de capacidades lógico/argumentativas. En segundo lugar, se impone el examen de la distinción entre *pensar* y *pensar críticamente*. Al respecto, conviene recordar que no siempre, ni en todas las circunstancias, el ejercicio de la facultad de pensar puede ser calificado como pensar crítico, ya que este se caracteriza por trascender y ubicarse fuera de los esquemas predeterminados, lo que origina una ruptura con la que se alcanza la autonomía y la libertad de acción.

Se propone, de manera provisional, adecuada a la reflexión propuesta, esta acepción de *pensamiento crítico*: ejercicio pleno de la facultad racional a partir de un proceso de ruptura con los esquemas dados, lo cual da como resultado la producción de nuevos conceptos los cuales han de ser sometidos reiteradamente al mismo procedimiento[4].

Ahora bien, más que definir, pareciera mucho más fructífero para propósitos académicos y formativos, consensuar qué no es pensamiento crítico. En este sentido, queda claro que el pensamiento crítico no es una ideología, tampoco es una técnica argumentativa, aun cuando la adquisición y el desarrollo de estas capacidades resultan fundamentales. No es una filosofía, pues no está comprometida con determinada ontología ni metafísica específica. Tampoco es producto de una postura pseudointelectual o academicista. En última instancia "pensar críticamente" es más una "actitud" que una aptitud; es una predisposición, un estilo de vida. Así, se puede afirmar que pensamiento crítico se contrapone a pensamiento único, pensamiento unilateral, pensamiento dogmático, pensamiento fundamentalista.

Pensamiento crítico: contextos referenciales

La reconstrucción racional del recorrido histórico-cultural que posibilita ubicar los hitos trascendentales para comprender el proceso evolutivo del pensamiento crítico se descubre como tarea ineludible, aunque en extremo compleja. Esta reconstrucción encuentra su momento fundacional en el paradigmático caso de los griegos al referirnos al proceso que describe el paso del mito a la razón (logos), también al momento presocrático y la distinción doxa/episteme.

El concepto pensamiento crítico reaparece en la temprana Época Moderna con el nuevo racionalismo y la ciencia moderna, con las egregias figuras de Francis Bacon[5], Galileo Galilei, Isaac Newton y, principalmente, Rene Descartes[6], considerado el "Padre del racionalismo moderno".

Se trata del señorío de la razón, fortalecida por el valor del testimonio experimental. Esta cultura racional, esta nueva modernidad, se manifiesta como ruptura y, posteriormente, como proyecto. Como ruptura, expresa la rebelión contra el dogmatismo imperante encarnado en el poder ideológico de la Iglesia, el aristotelismo y el escolasticismo: nace la ciencia experimental como clara ruptura ante la concepción clásica de ciencia entendida, cuya visión del mundo ~~que~~, aunque racional, se mantiene claramente en el plano especulativo. Asistimos así, al proceso inicial de la reivindicación de la razón frente a la fe, de la ciencia experimental frente a la Teología y del nuevo criterio de autoridad: el dictamen de la razón y la experiencia frente al dogmatismo encarnado en el aristotelismo y los textos sagrados.

Como proyecto, la nueva modernidad hace alusión al *proyecto ilustrado*: la razón científica como garantía de emancipación del hombre. Expresión, por antonomasia, cuyo espíritu ilustrado es Immanuel Kant, quien propugna por la necesidad de superar el pensamiento dogmático y asumir la filosofía crítica.[8] Posteriormente, Karl Marx sistematizaría una crítica de la economía política, la filosofía y la religión.

Ya en el siglo XX, y lo que transcurre del presente, el movimiento neopragmatista propugna por el abandono de las "filosofías de la conciencia"[9]; y Popper (1974) propone entonces el *racionalismo crítico* en el contexto de una epistemología sin sujeto cognoscente.

Desde la Teoría Social, la Escuela de Frankfurt da fundamento a una nueva reflexión acerca del orden social: la Teoría Crítica de la Sociedad[10].

Conviene recordar el mandato kantiano sobre asumir una actitud crítica frente a los tutores y los poderes instituidos: "ten el valor de servirte de tu propia razón *sapere aude*" (Kant, 1993)[11]. Desde la década de 1920, un grupo de intelectuales se reunía periódicamente en lo que posteriormente se conocería como el Instituto de Investigación Social de Frankfurt. De estas reuniones, y a partir de la exigencia de una relectura del marxismo, se consolida una crítica a la instrumentalización de la razón ejercida por la

Modernidad; sin embargo, no será ~~sino~~ hasta la década de 1940 cuando hace su aparición la Teoría Critica de la Sociedad. Esta corriente se designó en adelante como la Escuela de Frankfort; y entre sus pensadores más connotados y, explícitamente, considerados como frankfurtianos se citan a: Max Horkheimer, Theodor W. Adorno, Herbert Marcuse y Jürgen Habermas.

El desarrollo del pensamiento crítico

La actitud (auto)crítica y las competencias lógico/argumentativas

Propiciar la promoción del *pensamiento crítico* exige identificar y examinar las condiciones de posibilidad, tanto de naturaleza subjetiva como objetiva, del cultivo de este. En este sentido, referirse al valor del desarrollo del pensamiento crítico, implica promover un contexto cultural favorable para su construcción y adquisición; es decir, entender procesos de *aprender a desaprender* que posibiliten la reestructuración de nuestros esquemas conceptuales, culturales, metodológicos, actitudinales, ideológicos y axiológicos, en función de los cuales nos orientamos en el mundo. Esto implica la capacidad autocrítica como condición *sine qua non* para el desarrollo del pensamiento crítico.

Se trata de enfrentarnos al compromiso moral de desmontar, siempre y en cualquier circunstancia, toda forma de discurso falaz, sesgos y hábitos mentales, pero particularmente cultivar la capacidad autocrítica. Solo a partir de esto, se adquieren y desarrollan los procesos de adquisición de competencias lógico/argumentativas; toda vez que las personas permanentemente se enfrentan a dilemas que exigen tomar decisiones, algunas de carácter personal, otras de índole colectivo.

El proceso de toma de decisiones implica una estructura racional en la que las emociones desempeñan, en ese proceso, un importante papel; entonces, se trata de aprender a conjugar ambas estructuras. Pensar lógicamente, en contextos de decisión, significa pensar de manera eficaz, lo cual supone obtener lo que nos hemos propuesto en el momento cuando lo planeamos; pero también implica, pensar de manera eficiente, es decir, que nuestra reflexión nos guíe al mejor aprovechamiento de los recursos disponibles (materiales, económicos, cognitivos, sociales, humanos, etc.) a fin de alcanzar el objetivo propuesto. Por otra parte, una decisión eficiente es aquella que no genera más problemas de los que resuelve. Se impone así, la necesidad de buscar un equilibrio entre eficiencia y eficacia.

El concepto de *racionalidad* debe incorporar pues, una connotación de racionalidad ética que posibilite la toma de decisiones que, además de eficaz y eficiente, sea ética, esto es que traiga el mayor beneficio a los involucrados, evite generar daño, y si fuera inevitable, el menos posible.

La racionalidad de la decisión no radica en la decisión tomada, sino en el proceso lógico de análisis que llevó a esta.

La decisión –solución a nuestro problema– no es, en muchas ocasiones, lo más difícil, sino la voluntad para realizarla. En ese caso, la realización de la decisión se convierte en el problema, más que la toma de la decisión en sí.

Aun cuando la lógica no se vincule directamente con nuestras emociones, en cierto sentido tampoco está desligada, dado que tomar una decisión emocionalmente adecuada puede suponer un buen razonamiento que la respalde.

La ayuda que provee la lógica no se limita al ámbito personal, pues, por la presencia de otras personas con las que convivimos y nos relacionamos, requerimos justificar nuestras acciones, basarlas en razones objetivas y compartibles por todos, ya que necesitamos acuerdos que nos permitan alcanzar una vida armónica y justa.

Así pues, la adquisición y desarrollo de competencias lógico/ argumentativas nos capacitan para reconocer argumentos; valorarlos como válidos o inválidos; distinguir entre premisas y conclusión; construir argumentos de cara a la defensa de un punto de vista; reconocer argumentos falaces, y dominar técnicas para desmontar dichos argumentos falaces[12], identificar las diversas formas de razonamiento lógico para aplicarlas en procesos de redacción y comunicación oral; argumentar en favor de posiciones teóricas propias de cara a su aceptación por los demás; y sobre todo hacerles frente a las trampas del lenguaje demagógico y sofistico.

Niveles de aplicabilidad

Este aspecto apunta tanto al examen crítico de los diversos campos disciplinarios como al nivel de aplicabilidad del pensar crítico, fundamentalmente en torno al orden político/social, y cultural, por una parte, así como a la dimensión de la propia vida personal por otra (Marinoff, 2000; Ruiz, 2018).

Si existe un ámbito entre las disciplinas con pretensiones de cientificidad, en el cual la ausencia del pensamiento crítico se manifiesta de manera clara y evidente, lo podemos ubicar e identificar, sin duda alguna, en ciertas corrientes, escuelas y prácticas, a lo interno de las llamadas ciencias sociales, caracterizadas por el papel cuasi determinante que desempeña el factor ideológico en el desarrollo de sus marcos heurísticos.

Hablemos entonces del pensamiento crítico en el ámbito de las ciencias sociales y las humanidades como lo son el análisis crítico de la economía, de la sociología, de la politología, de la antropología, de las ciencias jurídicas, y demás áreas de conocimiento (Bunge, 1999). Igualmente, en campos como la educación, el arte y la religión.

El pensamiento crítico: democracia, ciudadanía y el orden político

Como colofón del presente escrito, conviene referirse al tema de la relación entre *pensamiento crítico* –tal como lo hemos tratado aquí–, la *democracia* y el *orden político*, entendidos estos como el ámbito y la forma de convivencia entre las personas.

Es fácil advertir que, en nuestro medio, tanto la *democracia* como la *política* se han desnaturalizado, han desvirtuado su sentido, naturaleza y razón de ser. En consecuencia, la *democracia* y la *política* se han constituido en rehenes de los más oscuros intereses mercantiles; a la par, sufren la grave patología de la lumpenización, lo que divide a la democracia, en deformaciones manifestadas en sus dos expresiones: la *demagogia* y el *democratismo*. La *política*, por su parte, ha sido vaciada de su contenido ético/moral, se ha desencajado de su original sentido y finalidad: propiciar el bien común.

Sin duda alguna, el más noble invento humano es la *política*, no la ciencia ni la práctica, sino esa forma específica de convivencia entre los seres humanos que expresa la constatación del triunfo de la razón sobre la barbarie, la concreción del hombre en sí y para sí, el ámbito de conciliación entre naturaleza y libertad a la manera en que la pensó el filósofo ilustrado por excelencia: Immanuel Kant.

No obstante, sin ciudadanos que piensen críticamente, la *democracia* y la *política* resultarían vacías, estériles, disfuncionales e ineficaces. Por ende, el ejercicio de la ciudadanía es impensable sin el pensamiento crítico; este es impensable en regímenes caracterizados por la ausencia del Estado de derecho y una cultura de respeto a los derechos humanos, el desprecio a la dignidad de la persona humana, y una ética que atenta contra el principio kantiano de no tratar nunca a la persona como medio, sino siempre como fin. (Kant, 1974)

La lucha por el control de las conciencias mediante técnicas heterogéneas estructuradas como redes de poder, a las que Foucault (1993) denominó "tecnologías del poder" (citado por Ayestarán Uriz, 1996), buscan manipular y determinar la conducta de los individuos sometiéndolos a ciertos tipos de fines, siempre en detrimento de sí mismo mediante el uso de medios de comunicación masiva, de sistemas educativos alienantes, y del comercio con la necesidad espiritual de las personas. Todo esto aunado a la mercantilización y lumpenización de la política, a la promoción y tolerancia de la seudociencia y del fanatismo religioso que atentan permanentemente –de manera intencional– contra el desarrollo del *pensamiento crítico* (Dawkins, 2017). De ahí el compromiso de instituciones como la Universidad de Panamá de consagrar y promover el cultivo del pensamiento crítico en sus normas, estatutos, reglamentos y discursos, además de combatir y rechazar, constante y permanentemente, todo lo que, continuada y sistemáticamente, atenta contra esa misión.

Un camino para alcanzar esa misión es priorizar, como un imperativo ético/moral la educación infantil, ya que aprender a pensar críticamente desde pequeño propicia la práctica de someter a análisis las propias ideas, creencias, deseos, presuposiciones.

Los niños deben aprender a pensar de manera independiente, a identificar y evaluar argumentos, y a distinguir entre hechos y opiniones. Y la filosofía resulta un instrumento efectivo en este propósito.[13] Lo siguiente sería un sistema educativo enfocado en ensenarles a niños, jóvenes y adultos, a plantear(se) preguntas adecuadas, exigir respuestas, y saber cribarlas a la luz del pensamiento crítico.

La *democracia*, como el menos malo de los sistemas políticos, requiere permanentemente someterse a procesos de *asimilación* y *acomodación*[14] a fin de asegurar su funcionabilidad vigencia y legitimidad. Para garantizar esto, es imprescindible un contexto cultural favorable que fomente el *pensamiento crítico*. Este proceso debe iniciar desde las primeras etapas de la vida del individuo.

Lo anterior requeriría, por supuesto, una revolución en el ámbito de la escuela, ~~un~~ el cambio del viejo paradigma a otro fundado en tres pilares: la enseñanza práctica de las ciencias y el razonamiento lógico; la exclusión total de la religión (por su nefasto y pernicioso efecto enajenante en la mente de los niños y jóvenes); y el enfoque, no tanto de qué pensar, sino de cómo pensar. Solo así nos constituiríamos en adultos libres de dogmas, de fanatismos y fundamentalismos (religiosos, ideológicos); con capacidad para evaluar, cuestionar, ponderar de manera crítica, racional y autónoma, en todo contexto y circunstancias. Y, en consecuencia, las acciones de nuestros gobernantes, serían garantes y defensoras de una sociedad auténticamente democrática.

Conclusión

Democracia y *pensamiento crítico* son dos caras de la misma moneda. Sin un *pensamiento crítico* robusto, la democracia se debilita, se vuelve susceptible a las manipulaciones y pierde su capacidad de servir como un verdadero reflejo de la voluntad popular. Por ello, su promoción debe ser una prioridad para todas las instituciones educativas y culturales, y así coadyuvar a mantener vigente una sociedad libre, justa y equitativa.

Notas

[1] El término “critica” en su sentido etimológico del griego *krises*: separación, escisión; por extensión elección, resolución, desenlace. El verbo correspondiente *krinein* alude a discernir, separar, escoger, decidir. Para los fines de este ensayo, entenderemos *critico* en el sentido de ruptura; y por extensión, ruptura con lo dado.

[2] *Sin desmeritar la rica tradición de la sabiduría oriental, privilegiamos la concepción según la cual el pensamiento* filosófico propiamente dicho surge como una actividad explícitamente preocupada por establecer sistemas conceptuales, fundados lógicamente, y con pretensiones de verdades razonadas y aceptables precisamente en virtud de la lógica interna de sus propios enunciados. En consecuencia, la razón, es decir el ejercicio de la razón, vendría a constituirse, tanto en el ámbito como, en el método del pensar filosófico, complementado esto con las exigencias de un lenguaje objetivo, divorciado de emotividades psicológicas y/o de la sensibilidad religiosa.

[3] La expresión filosofía occidental resultaría tautológica toda vez que “la Filosofía es, en esencia, griega...la Filosofía ha recurrido, en primer lugar, a lo griego –y sólo a ello– para desarrollarse” (Heidegger, 1956/1978).

[4] Esta acepción provisional implica por supuesto, el examen de una Teoría de la Racionalidad, pero esta tarea excede los objetivos de este ensayo. En términos generales se entiende la racionalidad como el ejercicio de la capacidad llamada razón. En función de esta capacidad, le es permitido a quien la ejerce (el agente racional) tomar decisiones y llevar a cabo acciones para manejarse con su realidad, bregar con sus condiciones, emociones, y circunstancias, es decir, sobrevivir. Esta capacidad está, a su vez, compuesta de otras capacidades básicas tales como: tener representaciones del mundo; hacer conexiones entre representaciones; tener creencias; proponerse fines que alcanzar; conectar unas creencias con otras; elegir en ciertas circunstancias entre cursos de acción; asociar representaciones con términos lingüísticos;

aprender y usar un lenguaje proposicional; expresar creencias por medios de proposiciones; conectar algunas proposiciones con otras y hacer inferencias; comunicarse con otros agentes racionales; dialogar, dar y recibir razones con otros agentes racionales; construir modelos para proponer guías de acción; elegir en ciertas circunstancias entre modelos; tener por valiosos ciertos objetos y estados de cosas en el mundo; evaluar, en ciertas circunstancias, determinadas creencias; justificar inferencias, creencias, y cursos de acción; evaluar en ciertas circunstancias los fines propuestos, sus intereses, principios, y valores con el fin de mantenerlos, modificarlos, o cambiarlos. Véase Olive (2011).

[5]Junto con Galileo Galilei, Francis Bacon es considerado uno los gestores de la ciencia moderna. Véase O'Connor (1985). Su aporte al desarrollo de la nueva racionalidad es inestimable. Considera que el camino de la inducción es el método correcto, no obstante la correcta utilización del mismo exige, como condición previa, superar, la inclinación natural hacia el error; por eso, antes de enseñar y aplicar dicho método, hay que superar los prejuicios que dominan la mente humana*;* estos prejuicios, a los cuales Bacon denomina *ídolos,* se presentan como nociones o ideas falsas que se apoderan de la mente y tienden siempre a reaparecer, estas pueden ser innatas o adquirida y son categorizadas por Bacon de la siguiente manera:
Ídolos de la tribu: Son comunes a todos los seres humanos; se fundamentan en la naturaleza humana. Aluden a la tendencia a suponer un orden en la naturaleza diferente al que realmente existe. ejem. asignar a los astros orbitas circulares perfectas; la tendencia a generalizar las opiniones; tendencia a conferir fines y metas a los fenómenos naturales.

Ídolos de la caverna: Son propios de cada individuo, el cual es como una caverna en la cual se deforma la luz de la naturaleza; provienen del temperamento, la educación, la lectura, las experiencias particulares, las costumbres, la influencia de la autoridad de las personas que admiramos, la cultura, etc.

Ídolos del foro:(la plaza pública): Proceden de la relación social entre los hombres, y radican en la fuerza de las palabras que trasmiten nociones fantásticas y perturban las mentes. A juicio de Bacon son los ídolos más peligrosos toda vez que las palabras aparecen como sustitutos de la realidad; los hombres se ven lanzados por las palabras a innumerables disputas. La mayor parte de las controversias versan sobre palabras y no sobre la realidad de las cosas.

Ídolos del teatro: Proceden de los sistemas filosóficos anteriores; de sus métodos y lógicas, todos los cuales son como "*mundos ficticios y teatrales*" Ejemplo. los principios y axiomas de las ciencias que siguen prevaleciendo gracias a la tradición, la credulidad, la negligencia, y las malas reglas de la demostración. Por eso, el método propuesto por Bacon pretende luchar contra un poderoso enemigo que está en la misma mente: el ídolo de la lógica vulgar, y de la lógica aristotélica.

[6] El gran aporte de Descartes al desarrollo de la ciencia moderna, y en esto prima cierto consenso, lo constituye el valor de la duda metódica, expresión paradigmática de la rebelión contra la dominante visión clásica asentada en el escolasticismo, el aristotelismo, y el dictamen de las doctrinas de religiosas.

[7] La aparición de la ciencia moderna sólo puede ser comprendida si se abordar como un componte y expresión de la Modernidad, entendida, esta, como un complejo proceso socio/cultural que se manifiesta, por una parte, como ruptura, y por otra como proyecto. En efecto, se trata de una ruptura en todo sentido, y con manifestaciones culturales amplia y diversa, cuyos efectos se dejan sentir en un ancho espectro de la dinámica y estructura de la sociedad.

[8] Es justo reconocer aquí la labor titánica, enmarcada en aquella lucha encarnizada contra el poder político/teológico, llevada a cabo por científicos y filósofos durante el período conocido como Renacimiento. Resaltan en este sentido Nicolás Copérnico, Johannes Kepler, Giordano Bruno (quemado en la hoguera por orden del poder eclesiástico).

[9] Conviene aquí recordar el mandato kantiano de asumir la actitud crítica frente a los tutores y los poderes instituidos: "ten el valor de servirte de tu propia razón" "*sapere aude*". (Kant, 1993) Los tutores de los que habla Kant son aquellas personas, instituciones sociales, educacionales, políticas, religiosas que nos dicen que es mejor hacer lo que ellos nos advierten. Estos tutores tienen a su disposición una serie de recursos y estrategias para evitar que estemos en capacidad de activar nuestro mecanismo de pensamiento crítico. Nos insuflan el miedo de las consecuencias que puede acarrearnos el tener autonomía de pensamiento. Usan mecanismo de control mediante los cuales imponen su autoridad de manera sutil bajo el papel de protectores de la humanidad, y aprovechan estas circunstancias para inculcar a sus víctimas el germen del miedo.

[10] La filosofía de la conciencia fue inaugurada por R. Descartes a partir del descubrimiento del *cogito ergo sum*.

[11] Desde la década de los 20 del siglo XX un grupo de intelectuales se reunían periódicamente en lo que posteriormente se conoce como el Instituto de investigación social de Frankfurt. Producto de estas reuniones y a partir de la exigencia de una relectura del marxismo, se consolida una crítica a la instrumentalización de la razón ejercida por la Modernidad; no obstante, no será sino hasta la década de los cuarenta del siglo XX cuando hace su aparición la teoría critica de la sociedad la cual se distinguió en adelante como la Escuela de Frankfort, entre cuyos autores más connotados y, explícitamente, considerados como frankfurtianos se citan a: Max Horkheimer, Theodor W. Adorno, Herbert Marcuse y Jürgen Habermas.

[12] La Lógica clásica distingue dos categorías de razonamientos falaces, a saber: aquellos cuya estructura presenta una ausencia de atinencia lógica de su(s) premisa(s) con respecto a su conclusión; son las llamadas falacias inatingentes; y aquellos estructurados en un lenguaje ambiguo, oscuro o vago, son llamadas falacias de ambigüedad.

Entre las falacias clásicas del primer tipo distinguimos: el *argumentum ad baculum*, (apelación a la fuerza); *argumentum ad hominem* (contra el hombre); *argumentum ad ignoratiaam* (afirmar la verdad de una tesis sobre la base de la no demostración de su falsedad); *argumentum ad miserircordiam* (recurso a la piedad); *argumentum ad populum* (llamado emocional a la multitud); *argumentum ad verecundiam* (apelación al sentimiento de respeto por personas famosas y/o de gran prestigio y autoridad); accidente (aplicar una regla general a un caso particular cuyas circunstancias hacen inaplicable la misma; o bien, derivar una regla general partir de un caso o pocos casos particulares).

Ejemplos del segundo tipo identificamos: el equívoco; la anfibología; división y composición; etc. Entre otros ejemplos de argumentos falaces tenemos: la causa falsa, la petición de principio, el hombre de paja, la pregunta compleja, la falacia booleana, la falacia democrática; la falacia genética. Para una ampliación de este tema, remitimos al lector a (Copi, 1985; Warburton, 2005; Gómez, 2010, Novella et al., 2020).

[13] Desde las últimas décadas del siglo XX, se viene promoviendo, en diversos países de Europa y Norteamérica una propuesta educativa conocida como filosofía para niños.

[14] Piaget introduce las nociones de *asimilación y acomodación* para dar cuenta de los dos aspectos del desarrollo intelectual el cual es presentado como un proceso de adaptación que prolonga la adaptación biológica. Aquí lo utilizamos en sentido metafórico asumiendo, metodológicamente, la democracia como un organismo (sistema) vivo.

Referencias bibliográficas

Ayestarán Uriz, I. (1996). Foucault y la muerte del hombre: Individuo y poder como tecnologías. En A. Alonso, I. Ayestarán Uriz y N. Ureña (Coords.), *Para comprender ciencia, tecnología y sociedad* (pp. 137-148). Editorial Verbo Divino.

Bunge, M. (1999). *Las ciencias sociales en discusión. Una perspectiva filosófica*. Editorial Suramericana.

Copi, I. M. (1985). *Introducción a la lógica*. Eudeba.

Dawkins, R. (2017). *El espejismo de Dios* (N. Pérez-Galdós, Trad.). Espasa.

Foucault, M. (1993). *Las redes del poder* (E. Díaz, Trad.). Editorial Almagesto.

Gómez, P. P. (2010). *Argumentación, teoría y práctica. Manual introductorio a las teorías de la argumentación*. Editorial Universidad del Valle.

Heidegger, M. (1978). *¿Qué es filosofía?* (J. L. Molinuevo, Trad.). Editorial Narcea.

Kant, I. (1993). Respuesta a la pregunta: ¿qué es la Ilustración? En I. Kant, *¿Qué es la Ilustración?* (pp. 17-25). Editorial Tecnos.

Kant, I. (1974). *Fundamentación para la metafísica de las costumbres* (trad. A. García Moreno). Edit. Nacional.

Marinoff, L. (2000). *Más Platón y menos Prozac. Filosofía para la vida cotidiana*. Ediciones B.

Mosterín, J. (1987). *Racionalidad y acción humana*. Editorial Alianza.

Novella, S., Novella, B., Santa María, C., Novella, J., & Bernstein, E. (2020). *Guía del universo para escépticos. Cómo saber lo que es real en un mundo cada vez más falso*. Océano.

O'Connor, D. J. (1985). *Historia crítica de la filosofía occidental*. Paidós.

Olive, L. (2011). La razón naturalizada y la racionalidad plural. En A. R. Pérez Ransanz y A. Velasco Gómez (Coords.), *Racionalidad en ciencia y tecnología* (pp. 29-38). Universidad Nacional Autónoma de México.

Popper, K. (1974). *Conocimiento objetivo* (trad. C. Solís Santos). Tecnos.

Ruiz, J. C. (2018). *El arte de pensar. Cómo los grandes filósofos pueden estimular nuestro pensamiento crítico*. Editorial Almuzara.

Warburton, N. (2005). *Pensar de la A a la Z*. Editorial Gedisa.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 182-195

Ensayo/Essay

Distinción y complementariedad entre ética y moral. Un enfoque desde la filosofía moral

Distinction and complementarity between ethics and morality. An approach from moral philosophy

Alessandro Caviglia

Pontificia Universidad Católica del Perú, Perú
a.caviglia@pucpág.pe
https://orcid.org/0000-0002-8106-7201

DOI https://doi.org/10.48204/2805-1815.8481

INFORMACIÓN DEL ARTÍCULO

Recibido el: 2/5/2025

Aceptado el: 21/8/2025

Keywords:

Ethics, morality, damaged ways of life, person, thing

Palabras clave:

Ética, moral, formas de vida dañada, persona, cosa.

ABSTRACT/RESUMEN

Abstract:

This paper addresses the distinction and complementarity between ethics and morality, seeking to clarify the philosophical implications of this case. It begins by examining philosophical approaches that view ethics and morality as synonymous terms, then presents ethics as a study of morality, understood as a synonym for the customs of a locality. Thus, Giusti, following Hegel, establishes the distinction between "morality" and "ethical life" to present two approaches for providing a rational foundation for customs. We point out, rather, that ethics aims to analyze ways of life, while morality introduces the negative principle of not confusing a person with an object, which allows for the correction of ways of life when they are damaged by phenomena such as domination or moral damage of various kinds.

Thus, ethics concerns the ends and values we project onto things, as well as the value we project onto personal and community relationships, such as friendship, romantic relationships, or family relationships. In such relationships, forms of abuse and domination by one party over the others can occur because some people may claim the right to unilaterally define and redefine the meaning of the relationship, generating a damaged way of life because some people are treated as mere means or objects. In this situation, it becomes necessary to introduce the moral principle, which is completely external to the ethical way of life, to repair the damage it generates. Only in this way would be damage caused be repaired, since ethics lacks sufficient resources to correct itself satisfactorily and clearly establish the distinction between people and things.

Resumen:

El presente ensayo aborda la distinción y la complementariedad entre la ética y la moral, con el objeto de establecer la aclaración filosófica del caso. Comienza con las aproximaciones filosóficas que entienden la ética y la moral como sinónimos a presentar a la ética como un estudio de la moral, entendida como análoga a las costumbres de una localidad. Así, Giusti, siguiendo a Hegel, establece la distinción entre “moralidad” y “eticidad” para presentar dos enfoques para otorgar un fundamento racional a las costumbres. Señalamos que la ética tiene como objetivo analizar las formas de vida, mientras que la moral introduce el principio negativo de no confundir a una persona con un objeto, posibilita corregir las formas de vida cuando estas se encuentran afectadas por daños morales de diversa índole como la dominación.

Así, la ética versa sobre los fines y valores que proyectamos a las cosas, así como el valor el valor que proyectamos a la relaciones personales y comunitarias, como es el caso de la amistad, las relaciones de pareja o las relaciones familiares. En esas relaciones, pueden producirse formas de abuso y dominación de una de las partes sobre las demás debido a que algunas personas pueden atribuirse el derecho de definir y redefinir unilateralmente el significado de la relación. Esto genera una forma de vida dañada, pues algunas personas son tratadas como meros medios u objetos. En esa situación, es necesario introducir el principio moral, ~~el~~ completamente externo de la forma de vida ética. Solo así se recompondría el daño que ello genere, ya que la eticidad carece de los recursos suficientes para corregirse a sí misma de manera satisfactoria y establecer con claridad la distinción entre personas y cosas.

Introducción

Es habitual utilizar los términos “ética” y “moral” indistintamente, según el contexto Giusti (2007) señala que “moral” es “el sistema de valores inmanente a una determinada comunidad, mientras que ‘Ética’ sería más bien la reflexión filosófica sobre el sentido de dichas normas morales” (pp. 19-20). En el terreno de la filosofía moral, se suele presentar ambos términos como sinónimos, hasta se afirma que ética procede del griego *ethos*/*ethiké* y que moral proviene de la voz latina *mos/moris*/*mores*, pero que ambos significan lo mismo. Gutsti aclara:

> Desde el punto de vista etimológico, los términos “moralidad” y “eticidad” son equivalentes. “Moralidad” viene del latín “*mos, moris*”, “eticidad” del griego “*ethos*”; tanto “*mos*” como “*ethos”* significan “costumbre”, “hábito”. “ética” y “moral” son nombres sinónimos derivados de aquel significado originario de “costumbres”; en cuanto disciplinas filosóficas, ambas se

> proponen brindar una fundamentación racional de dichas costumbres (1999, p. 179).

Entonces, moral y ética son sinónimos desde un punto de vista etimológico. Pero Giusti (1999) añade una distinción entre *ética* y "eticidad", así como entre *moral* y *moralidad*: mientras *ética* y *moral* refieren a costumbre y hábitos, moralidad y eticidad son dos disciplinas filosóficas que tienen como objetivo dotar de fundamentos racionales a las costumbres, aunque mediante estrategias diferentes. Lo que distinguiría a la eticidad de la moralidad son las estrategias para fundamentar la *ética* o la *moral* (utilizando estos términos indistintamente). Así, en tanto la *ética* y la *moral* se refieren a las costumbres efectivas, *eticidad* y *moralidad* serían caminos diferentes para fundamentar racionalmente dichas costumbres[1].

Esta forma de pensar la diferencia entre *ética* y *moral* asocia a esta última con la forma en la que Kant buscó dotar de fundamento racional a las costumbres, mientras que vincula la *ética* con la forma en la que Aristóteles abordó el mismo problema.

Uno de los grandes aportes de Hegel, filósofo a quien Giusti adjudica la distinción entre moralidad y eticidad, es, precisamente, rescatar el aporte central de Kant: la valoración de la libertad entendida como autonomía, pero articularla por los recursos conceptuales de Aristóteles y, en cierta medida, también los de Platón[2]. Al trasladar la cuestión de la distinción entre ética y moral a la que existe entre "eticidad" y "moralidad" en Hegel, el filósofo peruano erige una barrera entre el pensamiento de los antiguos pensadores, fundamento de las costumbres, cuyo representante más destacado es sin duda Aristóteles, y el de los modernos, entre los que se destaca Kant (Giusti, 1999, p. 179 y ss.).

La pregunta que surge es si Hegel tiene razón (o si debemos elegir entre Aristóteles y Kant), o ¿tenemos otras alternativas? Queda claro que el trabajo de Aristóteles es completamente diferente al de Kant. Pero ¿ambos tenían el mismo objetivo? El debate en filosofía práctica ha estado en gran medida inclinado por la posición de Hegel, pero ello no significa que su posición haya encauzado la discusión. De hecho, en el debate anglosajón, impregnado por la lectura de Aristóteles y mediada por el escepticismo de Hume, se suele usar mayoritariamente el término *ethics* y se deja de lado el de moral. En contraste, se estila interpretar las cuestiones de filosofía práctica a la luz de una concepción naturalizada del bien y del valor. Incluso filósofos de orientación kantiana han trabajado las cosas en esos términos; destacan los herederos de John Rawls, como Andrew Reath (1997, 2006), Barbara Herman (2021, 2022) y Christine Korsgaard (1996, 2000), quien sobresale por interpretar a Kant en términos del bien.

En lo que sigue, defenderé la idea de que Aristóteles y Kant no persiguieron el mismo objetivo y que se pueden calificar más como complementarios que como rivales. Para ello, propongo una interpretación plausible entre ética y moral, para conectar con una posible complementariedad y, de esta manera, extraer las consecuencias de esa vinculación.

La ética

La *ética* se centra en los bienes, los fines y valores que se articulan dentro de una forma de vida particular; está asociada a una particular teoría del valor, que, en vez de ser objetiva, como en la teoría económica, es subjetiva. Fuera de la perspectiva humana, no hay nada que tenga valor en el mundo. Desde ese mismo punto de vista, ni el mundo, ni la vida humana tienen sentido en sí mismos (Nagel, 1995), son las personas quienes proyectan el valor a las cosas, y cada una las valora de manera diferente. Lo mismo sucede con el sentido del mundo y de la vida. No existe un sentido *objetivo* o *absoluto* del mundo o la vida, pues cada uno proyecta un sentido al mundo y a la vida.[3] Esto es así porque cada persona establece una relación particular con los bienes materiales o no materiales.

La *ética* tiene que ver con los bienes, los fines y los valores personales. Cada persona se propone metas para su vida acordes con la forma como cada cual comprende su vida y su relación con el mundo. Por eso, la *ética* supone una comprensión hermenéutica de la vida personal que puede incluir una dimensión narrativa bajo la pregunta ¿cómo he llegado a ser lo que soy? y la pregunta prospectiva ¿quién soy y quien quiero ser? Así, la *ética* tiene una dimensión relativa a la propia identidad de la persona (Habermas, 2008).

De acuerdo con cómo una persona va articulando su identidad, asimismo proyecta su valoración sobre las cosas y genera una forma de vida. Es dentro de la forma de vida donde el individuo establece/crea vínculos con otros, y de estas relaciones surge siempre la modalidad personal del tipo *yo-tú* (Buber, 2020, p. 11 y ss.), en vez del tipo *yo-nosotros*. Es así, porque al interior de las formas de vida se priorizan las relaciones personales sobre las relaciones comunitarias[4]. Una persona puede entrar en relación con personas de otras comunidades; no obstante, enfocarse en el marco comunitario podría afectar la percepción de las diferentes transacciones interpersonales que dos personas pueden establecer. La psicología transaccional ha venido estudiando este tipo de relación (Berne,1984). Esto no quiere decir que la relación con la comunidad dentro de la forma de vida desaparezca o sea irrelevante, sino que la relación personal con otra persona particular u otras personas particulares (la amiga, la pareja, la familia) es lo que se encuentra en primer nivel.

En las relaciones personales, se valoran ciertos bienes compartidos, como la amistad o el amor de pareja. En esas relaciones se van definiendo y redefiniendo permanentemente esos bienes. Para ilustrar, entre los amigos nos hacemos bromas que no podríamos hacerlas a otras personas con las que no tenemos ese grado de confianza. Lo mismo sucede al interior de una relación de pareja o dentro de una familia. Lo importante es que las personas involucradas en la relación tengan la posibilidad de definir y redefinir el bien de la amistad o el amor de pareja. Pero cuando solamente la persona se abroga el derecho de definir y redefinir la relación unilateralmente se produce una *distorsión* o un mal en la relación, que se suele denominar *vida dañada* (Adorno, 2006).

Al interior de una relación de pareja afectada por el machismo, el varón se atribuye el derecho de definir y redefinir unilateral y permanentemente los términos de esa relación Se genera, entonces, una relación de dominación donde la voluntad de la mujer queda completamente anulada porque impera la voluntad del varón. En contraste, lo que diferencia a la dominación de la interferencia es que en la primera una persona (o un grupo de personas) tiene el control de la voluntad de otra persona (o de otro grupo de persona); en cambio, la interferencia bloquea de la esfera de acción a una persona (o un grupo de personas) (Pettit, 1999). Cuando se ejerce dominación en las relaciones personales y sociales se produce el fenómeno que, desde Theodor Adorno (2006) se conoce como *forma de vida dañada*.

El problema estriba en que la *ética*, al centrarse en la forma de vida y en extraer todos sus recursos de ella, no cuenta con la posibilidad de hacer nada cuando estas formas de vida se encuentran dañadas por la dominación y el abuso. Por ejemplo, el discurso ético señala que no se debe intervenir en las prácticas de las rondas campesinas andinas de capturar a los delincuentes y de propinarles castigo físico, que es siempre humillante.

Williams (1998) afirma que los criterios éticos los adquirimos a través del lenguaje que aprendemos en la sociedad y que somos capaces de reflexionar sobre ellos para modificarlos, pero sin salirse del propio terreno de las formas de vidas; de esta manera, se produce un proceso de ampliación de las simpatías vía la socialización. La pregunta es ¿de dónde salen los criterios de crítica de la propia forma de vida? La respuesta que ofrecen los defensores del discurso ético es que este cubre todos los aspectos de la vida práctica y, por lo tanto, esos criterios surgen de la propia forma de vida. Pero si la misma forma de vida se encuentra dañada, ¿puede esta ofrecer criterios de calidad para la crítica de sí misma?[5]

La moral

Si bien, dentro de sus formas de vida, las personas pueden llevar adelante procesos de reflexión y de transformación de las relaciones que establecen, los recursos de los que disponen son inmanentes a la misma forma de vida. La cuestión estriba en los límites y el alcance de la crítica de las formas de vida que se construyen con los recursos éticos, es decir, desde dentro de la forma de vida misma. La *ética* brinda tres caminos que conducen a esa forma de vida: a) a partir de principios que provienen de la tradición de la sociedad o la comunidad; b) partiendo de un ejercicio hermenéutico de reinterpretación de la forma de vida; y c) recurriendo a la crítica inmanente, tal como lo propone la Teoría Crítica de la Sociedad[6]. El problema de estos tres caminos es que se fundan exclusivamente en elementos éticos y no reconocen la diferencia entre los recursos de la *ética* y los de la *moral*.

Frente a estos tres enfoques críticos articulados con los recursos propios de la eticidad, la *moral* presenta una perspectiva completamente distinta. Kant (2018) coloca la *moral* como principio la distinción fundamental que existe entre una persona y una cosa.

> Los seres cuya existencia no descansa en nuestra voluntad, sino en la naturaleza, tienen sólo un valor relativo como medios siempre que sean seres irracionales, y

> por eso se llaman cosas; en cambio los seres racionales reciben el nombre de personas porque su naturaleza los destaca como fines en sí mismos, o sea, como algo que no cabe ser instrumentalizado simplemente como medio y restringe así cualquier arbitrio (al constituir un objeto de respeto) (p. 428).

Líneas más abajo, Kant establece el principio moral, bajo la forma de un imperativo práctico, en los siguientes términos: “Obra de tal modo que uses a la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como un fin y nunca simplemente como un medio” (2018, p. 429).

Tal distinción no proviene de las relaciones internas de alguna forma de vida en particular ni recurre a los recursos de la vida ética. Se funda, más bien, en la constatación que ofrece Kant, y que Forst (2014) retoma, de que el ser humano es un ser dotado de razón (y, en tal sentido, autónomo) y que, al mismo tiempo, forma parte de un mundo socialmente constituido. En tanto ser dotado de razón, el individuo es capaz de tomar distancia de las relaciones interpersonales y de la sociedad en la que se encuentra, para, por medio de la reflexión, percatarse de la distinción moral entre personas y cosas. A partir de esa distinción, surge el principio moral de no tratar nunca a una persona como a un objeto. Este principio es de carácter negativo y señala lo que no se puede hacer en ninguna circunstancia. Para expresar su exigencia en términos positivos, diremos que se debe tratar siempre a una persona como un ser digno, por ser un ser racional[7].

La complementariedad entre la ética y la moral

A partir del principio negativo, que es siempre negativo, no se puede articular una forma de vida; más bien, la *ética* supone una serie de bienes, fines y valores que da sentido a la vida de las personas. Dentro del terreno de la *ética*, la vida humana tiene su sentido y en él existe el conflicto trágico entre bienes. Pero, como dentro de las formas de vida se puede generar el fenómeno de la dominación, en el que se difumina la distinción entre personas y meras cosas para los agentes, es posible que las formas de relación interpersonales o sociales terminen generando daños morales. Así que, cuando se generan formas de vidas dañadas, es necesario introducir el principio moral para generar formas de relación libres de dominación. Una vez que se ha resarcido la forma de vida, el principio moral ha cumplido su trabajo y se puede volver a activar la vida ética. Sin embargo, esto no implica que, una vez que volvamos a la ubicación de las formas de vida, podamos abstraernos del principio moral. El principio moral debe permanecer como sustrato de la vida ética, a pesar de ser un elemento completamente distinto a ella.

Por otro lado, el principio moral se articula entre las personas que no comparten una comunidad y que comparten relaciones impersonales. El respeto incondicional que se le debe a toda persona por ser persona, independientemente del tipo de relación que tengamos con ella, se encuentra en la base de la *mora*l. Se trata del trato que le debemos a alguien con quien tenemos discontinuidad social. Ese mismo respeto debe encontrarse en la base de las relaciones interpersonales de quienes comparten una forma de vida y una comunidad. Entre amigos, son permisibles las bromas pesadas, contrario con quienes solo compartimos relaciones impersonales (por ejemplo, las personas que encontramos en el trabajo o a un cajero en el supermercado). Aunque también sucede que en las relaciones personales debemos mantener el respeto debido.

La distinción habermasiana entre el uso ético y moral de la razón

Habermas (1998) nos ofrece una distinción complementaria a la que hemos señalado entre *ética* y *moral* con el fin de esclarecer esa distinción, recurre a la pregunta "¿qué debo hacer?", ante un problema práctico. Imaginemos que, yendo al trabajo en bicicleta, de pronto se malogra la cadena. En ese momento, surge la pregunta ¿qué debo hacer? Ante esta situación, pueden surgir varias respuestas: llevar la bicicleta a un lugar donde la reparen; tomar un taxi para dejar la bicicleta en casa y de allí al trabajo; llamar al trabajo para comunicar el percance por el que llegaré tarde. Algunas de estas posibles acciones pueden ser elegidas a la vez o decidir por alguna alternativa. Como fuere, estamos ante un problema que podríamos llamar *técnico*, y que Habermas denomina *pragmático*. Lo propio de ese tipo problemas es que no invocan juicios de valor ni la distinción entre lo bueno y lo malo ni lo correcto de lo incorrecto. Así como cuando se nos desamarran los pasadores de los zapatos y volvemos a atarlos; esta clase de problemas carece de importancia ética o moral.

Ahora bien, si nos hallamos en una situación distinta, como la de sentir nostalgia de los almuerzos dominicales familiares celebrados en la niñez (o alguien extraña las fiestas patronales de su localidad), entonces nos encontramos en una situación diferente y la pregunta "¿qué debo hacer?" tiene un sentido diferente. Posibles respuestas a esta interrogante en particular serían "proponer a mis hermanos que nos reunamos un domingo para almorzar como cuando niños o buscar otra forma de hacer frente a la cuestión". El problema ahora ya no es técnico (o pragmático), sino ético. Cuando involucramos elementos como la nostalgia frente a cosas que se valoran, las relaciones personales al interior de una forma de vida son valiosas. En ese sentido, lo *ético* se acerca a lo *estético*[8]. Nos ubica en el terreno de la proyección del valor que atribuimos a determinadas relaciones personales e involucra ciertos tipos de sentimientos que son de la misma naturaleza que los sentimientos estéticos (sentimientos de lo bello y de lo sublime). "Regresar al pueblo donde crecí, a la casa de mi niñez y el encontrarme con mis hermanos y familiares hace que se sienta *cálido por dentro*". Las prácticas religiosas se ubican también en el terreno de lo ético-estético, al igual que las prácticas culturales.

El ámbito de la *moral* es diferente: está hecha de otra tela. Imaginemos que me encuentro en una relación de pareja y que hemos proyectado un futuro juntos sobre la base de un compromiso asumido libremente. Sin embargo, por cuestiones laborales, debo viajar, y durante ese viaje encuentro a otra persona que me hace sentir enamorado. En este caso, surge nuevamente la pregunta ¿qué debo hacer? Y otras interrogantes como ¿debo cultivar una relación paralela sin que mi pareja lo sepa?, ¿debo ser sincero con mi pareja y contarle lo que me está sucediendo? y ¿debo combatir ese sentimiento y honrar mi compromiso? Lo que está en juego aquí es la calidad del trato que se debe tener con la otra persona, es decir, *la calidad de la relación*. No se trata de si se valora o no la relación, como en el caso de la ética. Lo que está en juego es el respeto debido a la otra persona. En este caso concreto, se puede decidir tratar a las personas como instrumentos para complacencia propia o tratarlas como seres dignos de respeto. El sentido de la exigencia que se encuentra en la pregunta ¿qué debo hacer? es claro en el caso de la moral: "debo tratar a las personas con respeto, como seres dignos, y no como simples instrumentos o cosas". En este terreno, debo tomar distancia de las consideraciones éticas para hacer valer el respeto debido a cada persona por el simple hecho de ser persona.

Con esto, Habermas (1998) nos ha ayudado a afinar un poco más la diferencia entre la *ética* y la *moral*. Mientras que la *ética* se centra en lo que es valioso para una persona; la *moral* nos ubica frente a los demás mediante la exigencia del respeto debido. Cuando nos encontramos en el terreno moral, el trato de respeto que debemos a cualquier individuo no se basa en lo que la otra persona haga o en la calidad de su conducta[9].

Consideraciones finales

En el presente ensayo, he presentado una interpretación alternativa de la distinción entre *ética* y *moral* que se suele usar en filosofía moral. Esta interpretación plantea la diferencia entre ambos elementos. La ética tiene que ver con la pregunta "¿qué hace que mi vida sea valiosa para mí?" y la pregunta moral es "¿qué tipo de trato nos debemos unos a otros?" Mientras que la ética se circunscribe a una forma de vida compartida con otros; la moral nos presenta una exigencia de trato de respeto incondicional hacia cualquier persona, ya sea que comparta con la misma comunidad o carezca de continuidad social con ella.

Queda claro que la exigencia moral se encuentra en un estrato fundamental de las relaciones prácticas. Este estrato moral es de índole completamente distinto a las exigencias éticas, y que estas deben encontrar en la moral un límite que no se puede pasar. Por más valioso que una forma de vida pueda resultar para las personas, si las relaciones que se establecen entre ellas están basadas en la dominación y el abuso, deben ser criticadas desde el terreno moral. Esto es así incluso si las personas involucradas acepten este tipo de relaciones porque se encuentran insertas en un relato ideológico que les impide preguntar por la justificación del tipo de relación que tienen. La crítica moral es clara en el sentido de que esa clase de relaciones debe cuestionarse. La manera como se deben corregir las relaciones ya se encuentra en el terreno político-técnico.

Una de las críticas habituales hechas a la moral (o a lo que podemos llamar también *punto de vista moral*) es que no dispone de herramientas para enfrentar el conflicto entre bienes (Williams, 2002) presentes en las tragedias griegas y que, sin duda, forman parte de la vida misma. Esta crítica se levanta directamente contra Kant, quien afirma claramente que no existe conflicto entre leyes morales. Así, en la *Metafísica de las costumbres* el filósofo de la Ilustración sentencia con claridad que "es totalmente impensable una colisión de deberes y obligaciones" (2005, p. 224). Crítica que resulta acertada si las leyes morales, o los deberes y obligaciones que Kant presenta se entienden como bienes. Pero hacer eso sería entender el terreno de lo moral como una dimensión de la ética. Esta absorción de la moral por el discurso ético es un error. Como hemos visto, si bien la ética y la moral se complementan a la hora de pensar la filosofía moral y pensar las relaciones prácticas entre los agentes, se debe mantener con claridad la distinción. Así que, si en el terreno de la ética existen conflictos entre bienes que son irreductibles y que debemos manejar en la vida, hay que tener claridad respecto a la naturaleza de la moral. La moral no es el terreno de la vida buena, los bienes, los fines y lo valioso, sino el terreno de la distinción entre personas y cosas.

En ese sentido, tampoco es cierta la afirmación de Hegel (2005) de que la *eticidad* y la *moralidad* son dos estrategias distintas de fundamentar racionalmente las costumbres. El filósofo idealista alemán señala que la *ética* [*eticidad*] y la *moral* [*moralidad*] tienen el mismo objetivo o la misma materia, a saber, la fundamentación de las costumbres. Pero ello no es así. Si bien la ética versa sobre las costumbres; la moral, sobre "las restricciones incondicionales" que debemos colocar a las costumbres. Como observamos, se trata de objetivos distintos, razón por la cual se debe recurrir a formas de argumentación diferentes para fundamentar sus propios objetos.

Finalmente, podríamos decir que la eticidad hegeliana tiene como objetivo fundamentar las costumbres mediante la introducción de una racionalidad dialéctica. En el caso de la moralidad, el objetivo es otro: fundamentar racionalmente la distinción entre personas y objetos, y con ello el tipo de exigencias morales que debemos hacer valer más allá de las costumbres que tengamos.

Notas

[1] Giusti recurre a la *Filosofía del derecho de Hegel* para justificar sus afirmaciones. Fue Hegel quien había establecido este "sistema de conceptos", pero como la posición de Hegel no ha zanjado este debate, he colocado las afirmaciones de Giusti en condicional.

[2] Hegel sospecha de la forma en la que Kant plantea el problema moral, pero sabe que no se puede regresar al mundo de la *polis* (es decir, las pequeñas comunidades políticas griegas de los siglos V y IV a. C.) poque su estructura interna distingue entre la misma la esfera pública o Estado (*polis*) y la casa o el hogar (*oikos*). En el mundo moderno en el que Hegel se encuentra (y que se inicia a partir del fin de la Edad Media) queda claro que va abriéndose paso un tercer elemento, a saber, la sociedad civil (o sociedad burguesa). Para Hegel, ese nuevo elemento, y las exigencias de la libertad, hacen que no se pueda volver al mundo clásico, sino que es necesario leer el mundo moderno con el lenguaje proveniente de Aristóteles.

[3] La idea de que pueda existir un sentido del mundo y de la vida humana objetivo que toda persona debería reconocer no solo es falsa, sino que es sumamente peligrosa. Quien declara que conoce ese supuesto sentido se cree que tiene el derecho de decirle a las personas cómo deben vivir, y hasta obligarlas a organizar su vida de acuerdo con dicho sentido. Esta tentación ha estado presente en ciertas versiones del cristianismo, o en ciertas lecturas del marxismo u otras ideologías. La estrategia más usual para defender este punto de vista falso ha sido la afirmar que cierta persona o grupo tienen un acceso

privilegiado a supuestos conocimientos metafísicos, como Dios, el alma humana, el mundo o la naturaleza humana. A veces se ha confundido el conocimiento que tiene la física sobre el cosmos con ese supuesto conocimiento metafísico, o el conocimiento que tienen los médicos del funcionamiento de ciertos sistemas del cuerpo humano (por ejemplo, el sistema respiratorio) con el de la naturaleza humana. La física y la medicina versan sobre fenómenos de conocimiento humano, los cuales no son objetos metafísicos que darían sentido a toda la realidad en su conjunto.

[4] Ciertamente, en los enfoques de orientación hegeliana, como los de Axel Honneth y Charles Taylor, las relaciones intersubjetivas (incluso las personales, por ejemplo: padre-hijo) preceden a la articulación del *yo*. En tales enfoques se defiende la idea es que nuestros conceptos éticos más básicos los adquirimos de nuestras prácticas personales y sociales tempranas. Pero el problema de tales enfoques es que todas las relaciones *Yo-Tú* se encuentran orientas y son leídas desde el prisma de la comunidad, en la que permanece como presupuesto incuestionable el que toda reflexión sobre los agentes humanos tiene que desembocar en un “nosotros comunitario” que supuestamente es constituyente del yo, su transfundo hermenéutico y su destino último. Tal forma de pensar se ha terminado constituyéndose en un discurso ideológico que no permite pensar adecuadamente los problemas prácticos.

[5] Se podría argumentar que la decisión a propósito de la fuente de la crítica nos enfrenta a la cuestión de si nos inclinamos o por una estrategia trascendente-constructivista o por una inmanente-reconstructivista, o si ambas pueden relacionarse. Agradezco a Ronald reyes por hacerme notar este asunto.

[6] La Teoría Crítica articula su crítica tomando como punto de partida los mismos principios que se encuentran en la sociedad, pero que se hayan contradichas por la realización efectiva de las relaciones sociales existentes.

[7] Desde John Stuart Mill (2005, p. 5), el utilitarismo ha tratado de fundar la dignidad de las personas en el hecho de ser “seres sintientes”, de tal forma que el concepto de dignidad abarca incluso a cierta clase de animales (que tengan sistema nervioso central y médula espinal). De esta manera, se ha buscado tener argumentos que fundamenten los derechos de los animales. La estrategia argumentativa se encuentra en que, si basamos la dignidad en la razón, no podríamos dotar de fundamento a los animales. Por ello, deberíamos basar la dignidad en algo que abarque a humanos como a los demás animales. La categoría de “seres sintientes” parece ser suficientemente abarcadora para fundar en ella la dignidad. Sin embargo, esta estrategia presenta problemas. ¿Por qué debemos cerrar el conjunto de seres dignos solo con aquellos que poseen cerebro y médula espinal? Otro problema más radical es que todavía es posible violar la dignidad de un ser sin ocasionarle ningún dolor. Imaginemos que un científico puede desconectar los nervios que permiten que un animal sienta dolor; en dicho caso, se podría rebanar su cuerpo de la manera que se quiera siempre que no se lo mate. En esta circunstancia, la idea de “ser sintiente” parece no servir para proteger la dignidad del animal. De este modo, parece necesario pasar a otro elemento que sea más adecuado para proteger la dignidad de los animales. Christine Korsgaard ha propuesto fundar la dignidad en el hecho de ser consciente. Su argumento señala que los animales tienen diferentes grados de conciencia y que lo característico de los humanos es que tenemos una “conciencia reflexiva” (2000, p. 124 y ss.; 2018, pp. 2 y ss.). Así, mientras que los demás animales tienen una conciencia “volcada hacia afuera”, los humanos tenemos la posibilidad de ser conscientes de nuestros estados de conciencia y convertirlos en objeto de nuestra reflexión. Dicha conciencia reflexiva es lo que Kant denominaba razón, y es por eso que se puede señalar que mientras que los humanos tenemos dignidad por ser seres racionales, los animales no humanos son dignos por el hecho de poseer algún grado de conciencia.

[8] La relación entre la ética y la estética ha sido explorada en algunos momentos. La base común es la “*aestesis*”, es decir, la sensibilidad. Proyectar valor sobre un objeto incluye la intervención de la sensibilidad, como en el caso de contemplar la belleza natural o la belleza de una obra de arte. En el ser humano, la sensibilidad no va desconectada de razones, pero las razones para valorar algo corresponden al terreno que Rawls denomina lo “no público”, mientras que las razones morales corresponden al terreno de lo público. Ambos tipos de razones se pueden fundamentar y criticar, pero son de naturaleza completamente distinta.

[9] La dignidad que se está respetando es inherente a la persona y no depende de su conducta, de forma que no podemos alegar que una persona que ha cometido un delito ha perdido su dignidad y no merece ser tratada con el mismo respeto con el que tratamos a otras. El argumento de que la dignidad de Gandhi es mayor que la de Hitler y que debe establecerse una distinción moral entre ambos no viene al caso. El que hagamos responsables a las personas por lo que hacen y juzguemos a los delincuentes por medio de las reglas del debido proceso es lo que corresponde, en virtud de su dignidad. Esto es así porque la dignidad no es sinónimo de impunidad.

Referencias bibliográficas

Adorno, T. W. (2006). *Minima moralia. Reflexiones desde la vida dañada*. Akal. (Obra original publicada en 1956)

Berne, E. (1984). *Games people play: The psychology of human relationships* [Los juegos que juega la gente: la psicología de las relaciones humanas]. Penguin Books. (Obra original publicada en 1964)

Buber, M. (2020). *Yo y tú*. Herder. (Obra original publicada en 1974).

Forst, R. (2014). *Justificación y crítica. Perspectivas de una teoría crítica de la justicia*. Katz Editores. (Obra original publicada en 2011).

Giusti, M. (1999). Moralidad o eticidad. Una vieja disputa filosófica. En M. Giusti, *Alas y raíces. Ensayos sobre ética y moralidad* (pp. 175-200). Fondo Editorial de la Pontificia Universidad Católica del Perú.

Giusti, M. (2007). Introducción: El sentido de la ética. En M. Giusti y F. Tubino (Eds.), *Debates de la ética contemporánea* (pp. 13-42). Fondo Editorial de la Pontificia Universidad Católica del Perú.

Habermas, J. (1998). *Escritos sobre moralidad y eticidad*. Paidós.

Habermas, J. (2008). *Facticidad y validez: Sobre el derecho y el Estado democrático de derecho en términos de teoría del discurso* (5ª ed.). Trotta.

Hegel, G. W. F. (2005). *Principios de la filosofía del derecho o derecho natural y ciencia política*. Edhasa.

Herman, B. (2021). *The moral habitat* [El hábito moral]. Oxford University Press.

Herman, B. (2022). *Kantian commitments: Essays on moral theory and practice* [Compromisos kantianos: ensayos sobre la teoría y práctica moral]. Oxford University Press.

Kant, I. (2005). *La metafísica de las costumbres*. Tecnos.

Kant, I. (2018). *Fundamentación para una metafísica de las costumbres*. Alianza Editorial.

Korsgaard, C. M. (1996). *Creating the kingdom of ends* [Creando el reino de los fines]. Cambridge University Press. https://doi.org/10.1017/CBO9781139174503

Korsgaard, C. M. (2000). *Las fuentes de la normatividad*. Universidad Nacional Autónoma de México.

Korsgaard, C. M. (2018). *Fellow creatures: Our obligations to the other animals* [Compañeros animals: Nuestras obligaciones hacia los demás animales]. Oxford University Press.

Mill, J. S. (2005). *El utilitarismo*. Alianza Editorial.

Nagel, T. (1995). *¿Qué significa todo esto?: Una brevísima introducción a la filosofía*. Fondo de Cultura Económica.

Pettit, P. (1999). *Republicanismo: Una teoría sobre la libertad y el gobierno*. Paidós.

Reath, A. (1997). Legislation for a realm of ends: The social dimension of autonomy [Legislación para un reino de fines: La dimensión moral de la autonomía]. En A. Reath, B. Herman y C. M. Korsgaard (Eds.), *Reclaiming the history of ethics: Essays for John Rawls* (pp. 214-239). Cambridge University Press.

Reath, A. (2006). *Agency and autonomy in Kant's moral theory* [Agencia y autonomía en la doctrina moral de Kant]. Clarendon Press.

Williams, B. (1998). *Introducción a la ética*. Cátedra.

Williams, B. (2002). *Moral luck: Philosophical papers 1973-1980* [Suerte moral: artículos filosóficos 1973-1980]. Cambridge University Press.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 196-211

Ensayo/Essay

From essence to construction: feminist revisions of identity

De la esencia a la construcción: revisiones feministas de la identidad

Aiswarya Pradeep Kumar

JSPM University, India
aiswaryapradeep25@gmail.com
https://orcid.org/0009-0002-4027-3145

DOI https://doi.org/10.48204/2805-1815.8482

INFORMACIÓN DEL ARTÍCULO

Recibido el: 31/07/2025

Aceptado el:

Keywords:

Self, subject, social construction, identity, gender performativity

Palabras clave:

Yo, sujeto, construcción social, identidad, performatividad de género

ABSTRACT/RESUMEN

Abstract:

This study traces the journey of the philosophical evolution of the “self” within the background of feminist theory across its three major waves. The study investigates how feminist thought has critically interrogated and transformed traditional conception of identity. The analysis is structured in three sections: The Autonomous self, From the Private to Political, Fragmented Selves and Fluid Identities. By drawing on key feminist thinkers such as Mary Wollstonecraft, Simone de Beauvoir, Judith Butler, this article shows how feminist discourse has shifted the notion of the self from a fixed, rational subject to a dynamic, socially constructed, and performative subject. The study concludes that feminist philosophy not only challenges essentialist views of gender but also redefines subjectivity itself, contributing to broader debates in contemporary political and philosophical thought.

Resumen:

Este estudio traza la evolución filosófica del "yo" en el contexto de la teoría feminista a lo largo de sus tres grandes olas. El estudio investiga cómo el pensamiento feminista ha cuestionado críticamente y transformado la concepción tradicional de la identidad. El análisis se estructura en tres secciones: El yo autónomo, De lo privado a lo político, Yoes fragmentados e Identidades fluidas. Basándose en pensadoras feministas clave como Mary Wollstonecraft, Simone de Beauvoir y Judith Butler, este artículo demuestra cómo el discurso feminista ha transformado la noción del yo de un sujeto fijo y racional a un sujeto dinámico, socialmente construido y performativo. El estudio concluye que la filosofía feminista no solo desafía las visiones esencialistas del género, sino que también redefine la subjetividad misma, contribuyendo a debates más amplios en el pensamiento político y filosófico contemporáneo.

Introduction

This study traces the formation of the self as a performative subject, constituted through lived experience and iterative engagement with social practices. The research critically examines the evolution of the self from the Enlightenment period to the postmodern age. The three sections of the paper critically analyse the concept of self, which is a performative construct rather than an innate or unchanging essence. This study offers a significant critique of the essentialist and universal framework of gender identity, as well as an opportunity to reconsider the ontology of gender identity as fluid and conditioned by political and social systems.

The autonomous self

Identity is an overarching concept in philosophical discourse. The concept of identity has evolved significantly from the modern era to the postmodern period, making a shift from the autonomous self to the postmodern subject. In the modern period of philosophy, the autonomous self is identified as stable, universal, and rational. In Cartesian philosophy, the notion of the 'self' is understood as asocial, atemporal, unchangeable and indubitable. Descartes's theory of the indubitable ego cogito is the culmination of the 'method of doubt' to arrive at the identity of the individual. Descartes' method puts everything into question including sense-testimony, truth of science, and existence itself. His search culminates into one thing, i.e., one cannot doubt one's own 'thinking.' Here, Descartes posits the theory of *Cogito ergo sum* (I think, therefore I am) to establish that it is a self-evident truth that can be known by reason. Descartes argues in Meditations that,

> I think, therefore I am, was so certain and so evident that all the most extravagant suppositions of the skeptics were not capable of shaking it, I judged that I could accept it without scruple as the first principle of the philosophy I was seeking. (Descartes, 1637/1968, p. 53-54)

Descartes's theory of the 'I' or the 'self' stands for a new paradigm in philosophy, prioritizing consciousness over the body. His contribution is novel in modern times and ushers in a new era to think about the status of human beings. Rationalists and empiricists thinkers approach the theory of the self in diverse ways. In this context, the contribution of Immanuel Kant holds significant importance in Western philosophy. Kant reconciles both empiricism and rationalism, positing that the mind has the innate idea of a ‘category of understanding’ to organize and structure empirical experience. Kant elaborates on the ethical view of the self in his work *What is Enlightenment.* In Kantian philosophy, the self is an autonomous agent, meaning an individual can make ethical decisions and exercise their own free will. For Kant, the existence of individuals is intertwined with reason. The autonomous self in Descartes is epistemic in nature which means it is the foundation of knowledge while Kant’s autonomous self-depicts self-legislation, especially in his moral philosophy.

The concept of the autonomous self, appeared with Descartes and Kant, severely influenced socio-political movements since then. Individual rights and autonomy laid the groundwork for the first-wave feminism, which started in the 19th century and the beginning of the 20th century. First-wave feminism advocated that, like men, women also should enjoy all privileges that exist in the society. Feminists consider the question of gender equality as primary and aim to create gender justice to ensure equal rights for women based on the idea of the equality of the sexes. First-wave feminism is a movement that advocates for equal rights for women, which came into prominence in the mid-19th century and early 20th centuries. The movement addressed the issue of disparity between men and women, particularly on the political and legal discrepancies that women faced.

Pioneering women’s rights figures such as Emmeline Pankhurst, Harriet Taylor Mill, Susan B. Anthony, and Mary Wollstonecraft, to name a few, largely addressed the issue of women’s suffrage, property ownership and educational access. The suffrage movement was a vital campaign that aimed to seek the right to vote for women, and this movement was considered as the underlying efforts of women who fight for the social and legal difficulties that disallowed them from taking part in a democratic process. Margaret Walters writes about the suffrage movement that

> The determination and the persistence with which women argued, and increasingly demonstrated, for the right to vote makes an inspiriting story; all the more so given

> the equal determination, and at times the virulence, with which their claims were opposed (Walters, 2005, p. 68).

In 1920, the United States passed the 19th amendment, granting the right to vote to women because of a result of tireless efforts from women.

The main aim of first-wave feminism was to abolish the societal constraints that restricted women's opportunities, and the movement advocated for other legal and social statuses of women, such as labor rights, marriage laws, and educational rights for women. From this movement, the feminists were challenged and deeply engaged in questioning patriarchal norms and discussing gender equality. Despite of this, the first-wave movement received criticism for primarily focusing on the issues of upper class and middle white class women.

Mary Wollstonecraft's contributions were a revelation in that era, and her ideas were mostly circulated through her controversial book *A Vindication of the Rights of Women*. The book rigorously questioned women's representation in society, which was always subordinate to men. To reach gender equality, Wollstonecraft proposed the importance of rational education in women's lives, arguing that the denying of education leads to their lives becoming miserable. In the second chapter, *The prevailing opinion of a sexual character* discussed, Wollstonecraft rebukes the social standards that are enjoyed by men, asserting that women are intellectually and morally different. She claims that both sexes have the right to practice for their own development. Wollstonecraft discusses how women are always in the state of ignorance due to their innocence, which she believes is a form of tyranny. She emphasizes the necessity of women to be recognized as rational individuals capable of enjoying virtue and happiness, rather than merely as objects of men's desire (Wollstonecraft, *1792/*1982, p. 19). In the book, Wollstonecraft famously asserts that "Taught from their infancy that beauty is woman's sceptre, the mind shapes itself to the body, and roaming round its gilt cage, only seeks to adorn its prison" (Wollstonecraft, *1792/*1982, p. 77). Wollstonecraft upholds that the identity of women is confined to their body rather than their intellectual capabilities. Women are restricted to developing their rational abilities due to the constraints of social norms surroundings their body. In the first wave movement, it can be understood that the individuality of women is aligned to the reason, which was inspired by the enlightenment period.

The renowned literary work *A Room of One's Own* by Virginia Woolf is a pioneering text in feminist literary criticism. Woolf expounds the importance of financial independence and personal boundaries in a woman's life. Woolf says "… a woman must have money and a room of her own" (Woolf, 1929/1977, p. 7) for her independence and creative freedom. Her work draws attention to the importance of freedom for artistic expression and intellectual space and points out the gender disparities that constrain economic self-sufficiency, women's education, and creative opportunities. By examining historical contexts, Woolf challenges the patriarchy that exists in literary tradition which excludes the recognition of women. Woolf argues, "Women have served all these centuries as looking-glasses possessing the magic and delicious power of reflecting the figure of man at twice its natural size. Without that power probably the earth would still be swamp and jungle. The glories of all our wars would be unknown" (Woolf, 1929/1977, p.41). Woolf analyzes the historical role assigned to women, which has been enhancing and supporting the self-image of men and their greatness. Men recognize themselves as greater than women through the validation received from them. The achievements of men are underpinned by the roles played by women in encouraging their confidence. This process always places women as subordinate to men. Woolf argues for the necessity to break free from this limitation and highlights the importance of rethinking the women's role in society. Woolf offers practical solutions for women's participation in cultural and literary life to overcome their material condition under patriarchy. Woolf believes that money and freedom are key to women's autonomy to demolish the sexual inequalities, and her work stays a powerful call for women to empower themselves in cultural and professional spheres.

The First-wave movement was fundamentally grounded in the principles of equality and justice. It was greatly influenced by the enlightenment era and liberal political philosophy. It marks a turning point, where the rise of establishing and advocating for a place for women in society is recognized as a significant step toward individual identity. In early times, individual identity was always associated with the category of men. It can be analyzed that the notion of 'self' or 'individuality' was always related to the context of rationality. Patriarchy was the leading factor contributing to the lack of independence in women, particularly through marriage, where women were denied the status of representing themselves as independent identities in a society.

The protests for suffrage, property rights, and access to education aimed to establish women as autonomous individuals with equal political and legal recognition, like men. From the background of the suffrage movement Susan B. Anthony argues, "…woman's need of the ballot, that she may hold in her own right hand the weapon of self-protection and self-defense" (Stanton et al., 1881, p. 22). This quote emphasizes that the right to vote, acts as a necessity for women to protect and defend their own interests. The power to vote fosters democratic engagement, giving women a voice in shaping the laws and policies that affect them. Here, the ballot is considered a powerful weapon for women's empowerment, serving as an indispensable aspect of their fight for freedom and equality. Thinkers like Wollstonecraft argue that, due to their exclusion from the public sphere, women are restricted from acting as independent agents in society. Individuality and reason should not be restricted to the category of men; they should be considered qualities that everyone can attain through rational thought, and they are not restricted to one gender.

From the private to the political

Second-wave feminism appeared in the 1960s and 1970s as an extension of first-wave feminism, expanding its aims beyond suffrage and property rights for women. Second-wave feminism addressed a broad range of issues such as gender role stereotypes, reproductive rights, workplace discrimination, domestic violence, and sexual liberation. Feminists primarily advocated access to abortion, contraception, equal pay in the workplace, laws against domestic violence, and challenged stereotypical gender roles within the family. Feminists sought to reconstruct the norms related to gender and sexuality. The slogan of second-wave feminism "the personal is political," signifies that a woman's experiences are not only personal but are also the result of larger social structures and power dynamics. For example, issues like domestic violence are not caused solely by the subjective experiences of a woman but are deeply rooted in gender inequalities within society. Through this, feminists aim to highlight how the experiences of women are profoundly influenced by social and political factors.

Body politics was a central idea in second-wave feminism that recognized the significance of women's autonomy over their body. The movement emphasizes raising awareness among women about their rights over their own bodies. It scrutinizes issues of objectification and violence against women, and reproductive rights which are beyond the purview of women and are controlled by men. In second-wave feminism, feminists primarily argued for bodily autonomy to establish the existence of women as having control over their own bodies. Feminists particularly focused on issues related to reproduction. At that time, women were often unable to make decisions about their own bodies, especially concerning pregnancy. The movement emphasized political engagement, encompassing campaigns advocating for women's access to contraceptives, the right to abortion, and the legislation of safe pregnancy termination. It successfully elevated knowledge about women's bodies, comprehensive sex education, and lobbied for maternal healthcare facilities to guarantee safe childbirth and postpartum care. Even though the significance of bodily autonomy had been identified and discussions on the notion of gender began, it emphasized that "Gender and second-wave feminism were born together"(Segal, 1999, p. 38). A prevailing idea that existed in second wave feminism was that the body is natural and gender is constructed.

The most celebrated work of this period was Simone de Beauvoir's *The Second Sex*, considered a foundational text of feminism that investigates the historical and cultural treatment of women. For de Beauvoir, femininity is a construction arrived through socialization, keeping male dominance. The book explores the historical and social contexts that subordinate women from men and examines how women are identified as passive, emotional, and nurturing to justify their marginalization.

The Second Sex mainly addresses the myths and stereotypes regarding the existence of women in society, de Beauvoir points out that “Reared by women within a feminine world, their normal destiny is marriage, which still means practically subordination to man; for masculine prestige is far from extinction, resting still upon solid economic and social foundations” (de Beauvoir, 1949/2011, p. 29). The actions of women in this feminine world are considered natural and related to biology. For de Beauvoir, femininity is a social construction rather than biological one; she posits, “One is not born, but rather becomes a woman” (de Beauvoir, 1949/2011, p. 273) which means that the traits and behaviors of being a woman are constructed and developed through social norms rather than inborn qualities. De Beauvoir emphasizes the process of ‘becoming’ and rejects the idea of essentialism. She posits that femininity is fluid and can vary across different historical contexts. The process of becoming a woman involves internalization of societal norms within a given social context.

This is exemplified through the gender roles that are culturally constructed. From the period of childhood itself, societal practices and cultural norms shape individuals based on their biological sex. For instance, within families, girls are often involved in activities associated with the notion of femininity, such as playing with dolls and adopting nurturing behaviors from their mothers. On the contrary, boys are engaged in activities related to physical strength. This process of socialization leads to the ‘becoming’ of an identity. De Beauvoir underscores the importance of cultural practices and institutionalized frameworks that shape an individual into a state of womanhood.

The process of ‘becoming’ a woman is deeply interconnected with the notion of the ‘other’ in de Beauvoir’s philosophy. In patriarchal society, women have historically been positioned as the ‘Other,’ while men are always identified with their own identity and individuality. De Beauvoir argues that, in legal contexts, the terms masculine and feminine are used symmetrically, but in practice, the relationship between the sexes is treated as unequal. Men are represented as universal and viewed with qualities of both positive and negative, but women are defined by specific limitations and portrayed with negative connotations. Men’s views are not considered questionable; on the other hand, women’s views are treated as invalid. Aristotle describes the state of women as lacking quality, and St. Thomas calls them an ‘imperfect man,’ an ‘incidental’ being. A woman’s identity is always related to her biological traits and often sees her as an object. Women are defined in relation to men; they do not have a status apart from a different sex. While women are considered secondary and aligned with sexuality, they exist only in relation to men (de Beauvoir, 1949/2011, p. 15). De Beauvoir says, “He is the Subject, he is the Absolute-she is the Other” (1949/2011, p. 16).

The notion of the ‘Other’ takes away women’s identity and independence. The identity of women is always acknowledged as a secondary position in society and subordinated to men. For de Beauvoir, the ‘Other’ is not a result of biological reality; rather, it is an effect of social norms and cultural frameworks. When de Beauvoir discusses women's freedom, she draws upon Sartre's notion of freedom. Sartre argues "To be free is to be condemned to be free" (Sartre, 1943/2003, p. 152). This quote emphasizes the dual nature of the state of freedom, which simultaneously acts as a privilege and as a burden. Sartre highlights that human beings are not constrained by a preordained essence or divine plan; rather, we create our essence through our actions, and freedom is inherent to human existence. The state of freedom is inescapable because it is a condition that is thrust upon human existence, and, in parallel, we cannot avoid the responsibility of our actions.

De Beauvoir was deeply influenced by the ideas of Sartre, particularly the notion of freedom. In a patriarchal society, women are assigned the role of the 'Other' as opposed to the absolute and superior powers that men enjoy. The process of 'othering' limits the possibilities of freedom and restricts women to the traditional roles of femininity. De Beauvoir wants to point out that for women to be absolutely free, they must break away from the social and cultural roles imposed upon them. De Beauvoir argues that a woman is inherently free and autonomous, just like any other person, but she is compelled to take on the role of the 'Other' under the patriarchal superiority. Here, a woman is labeled as an object and leads a life confined to a state of immanence by limiting her potential capabilities. Similarly, women, considered as 'Other,' do not transcend her aptitudes and remain curbed by the male superiority.

The 'immanence' in question is tied to a woman's body. For de Beauvoir, female embodiment is the objectified way of internalizing the gaze of others and constructing the body as an object for others. Women's bodies are constructed through societal influences from an incredibly youthful age. Girls learn about the societal meanings of "pretty" and "ugly" from cultural ideals presented through images and stories. They try to align with the beauty standards set in a society and adopt behaviors that seek validation. De Beauvoir wants to point out that the objectification of a woman's body is constructed through social norms and cultural frameworks. The female body becomes a site of oppression, influenced by societal pressures like gender roles, beauty standards, sexuality, and the tension between self-affirmation and societal pressure. For de Beauvoir, the body is represented as an experienced reality, which has its own identity and lived experience. De Beauvoir argues that "it is not the body-object described by biologists that actually exists, but the body as lived in by the subject" (de Beauvoir, 1949/2011, p. 65).

For Beauvoir, the body is not merely a physical entity; rather, it is shaped by societal taboos and laws. She argues that the subject is aligned with specific values and is not merely rooted in biological facts (de Beauvoir, 1949/2011, p. 63). When we try to understand the body of a woman, it is limited to biological facts, but de Beauvoir contends that "Woman is determined not by her hormones or by mysterious instincts, but by the manner in which her body and her relation to the world are modified through the action of others than herself "(de Beauvoir, 1949/2011, p. 681). As a result, a woman's experiences and capabilities are reduced into her physicality alone, and on the other hand a male body is viewed as active and free and stands to define meanings beyond its physicality. De Beauvoir does not see gender identity as an expression of biological sex, but rather as a cultural construction. She argued that gender identity is socially constructed. One is said to be masculine or feminine, not by any inborn essence in the body but on the other hand, by culture and history put together.

The Feminine Mystique, a pioneering work by Betty Friedan, played a significant role in the Second-wave movement. Friedan illustrates the problem with stories of severely unhappy stay-at-home spouses from the United States who struggled for fulfillment despite living in comfort and having seemingly "perfect" families. The book addressed the state of identity crisis of women in the role of a wife, mother, and homemaker, with women always acting according to the social expectations in their gender roles. Friedan coined the term 'feminine mystique' to idealize the image of femininity in the 1950s and early 1960s. It was also the post-World War II period where more restrictions were placed on women's activities, confining them largely to domestic chores and discouraging them from pursuing public education and careers. In the chapter *The Problem That Has No Name,* Friedan explains how the suburban stay-at-home spouses openly express their resentment towards merely fulfilling the duties of wives and mothers and confining themselves to the domestic chores. Most women of the time recognized from within a yearning for liberation. A liberation, not merely being a stay-at-home spouse but something beyond. Friedan explains it like this: "she was not talking about a problem with her husband, or her children, or her home. Suddenly they realized they all shared the same problem, the problem that has no name" (Friedan, 1963, p. 19). Friedan points out that women felt a quiet longing in their lives, seeking personal growth and identity beyond their traditional gender roles. Friedan highlights the relevance of recognizing the emotional and psychological needs of women, which were dismissed during that era.

Like Mary Wollstonecraft, Friedan emphasizes the significance of education as a vital tool for empowerment, and she argued that a well-rounded education was essential for women to achieve self-actualization and break free from the constraints of the 'feminine mystique.' By accepting and fulfilling these social expectations, women do not recognize their own identity, and a plethora of abilities are in store.

The most notable contribution of Second-wave feminism was that it was able to instill in women a notion of their own identity. It also opened debates on those social structures that shaped women's identity. While the first-wave movement focused more on acquiring legal and social rights, the second wave highlighted the cultural dimension of gender inequality. Feminists analyzed not only the explicit problems of women but also investigated the root causes of all their problems. From the slogan "the personal is political" to Friedan's statement "the problem that has no name," feminists emphasized the importance of equality and voiced against their subjugation based on gender. Both movements also realized that the existing social power structures contributed to it.

Another renowned American philosopher of feminist ethics and political philosophy, Marilyn Friedman argues that women have "suffered in many ways from social relationships, including the denial of whatever degree of personal autonomy might otherwise have been theirs (Friedman, 2000, p. 219)". She argues how the traditional conceptions of autonomy did not understand the impact of social relationships on individuals, especially on women. She emphasizes that women's freedom and personal autonomy have historically been constrained by various societal structures, and most of them result in various forms of suffering and limitation. The Second-Wave Feminist Movement questioned those prevailing patriarchal norms which considered women to be homemakers and caregivers. The significance of women's autonomy marks a significant turn in the second-wave movement. In that period, women gained access to reproductive rights, the legalization of abortion, and access to contraception, which until then was a monopoly in the hands of men. Women, who were confined to the private spheres of the family, began to fight for autonomy and individual freedom. The focus on individual liberation challenged the ideologies that constrained women's autonomy, and this moment placed a crucial step in the evolution of feminist philosophy.

The limitations of Second-Wave feminism led to the emergence of Third-Wave Feminism in the 1990s, which addressed the issues of diverse sets of voices and experiences. Sarah Gamble, a British academic and a feminist theorist argues that "Third wave feminism is characterized by a desire to redress economic and racial inequality as well as 'women's issues" (Budgeon, 2011, p. 310). The movement analyzes how a woman's experiences are shaped by several factors like class, sexuality, race, and nationality. It represents the voices of women from marginalized groups, and the Third-Wave movement continued to challenge the foundational assumptions of existing feminist frameworks that the culmination of a unified feminist subject is the result of that form of a feminism which focuses more on how the individual women understand and claim their own feminist identities (Budgeon, 2011, p. 9).

In earlier feminist movements, particularly in the Second-Wave, a unified experience of womanhood and their problems was analyzed, rather than drawing attention to the diversity of women's experiences. In the Second-Wave, though the issue was initially considered to be discrete problems of women from certain sections, in the later stage it was seen to be a widespread issue affecting a large group of women. And it should not be forgotten that these disparities were caused by the then existing societal structures. This awareness helped them to understand that personal struggles were no longer mere personal' but they were addressing much larger social issues of gender inequality (Snyder, 2008*, p. 184).*

Fragmented selves and fluid identities

Third-wave feminism advocated the freedom to express one's own identity, whether that includes embracing traditional gender roles or rejecting them, and challenged the societal norms around gender and sexuality. Feminists were aware that the individual issues vary significantly across distinct cultural contexts. Unlike the First-Wave and the Second-Wave, the Third-Wave movement witnessed the presence of various media platforms for activism. Third-Wave feminism critically analyzed traditional norms of gender and sex and tried to understand the intersectionality of individual experiences. Unlike other feminist movements, the Third-Wave focused on how gender and sex are influenced by intersecting factors like race, nationality, class, sexual orientation, and ethnicity. The primary concern of the movement was to deconstruct those binaries of gender and establish that gender and sex are socially constructed rather than biologically given entities.

Third-Wave feminism focuses on the fluidity of identities and the rejection of essential paradigms. Judith Butler is a prominent figure in the Third-Wave feminist movement. Her works contributed significantly to the Third-Wave movement as a critical lens to destabilizing the gender binary. Butler's *Gender Trouble: Feminism and the Subversion of Identity (*1990) is a revolutionary text in feminist and queer theory. Žižek considered *Gender Trouble* as a political practice and the anti-identitarian turn of queer politics (Žižek, 2000, p. 132). Butler claims to create and open a space for subjects who are marginalized by society because of denied recognition and the right to a livable life within existing social norms. With the saying "I want to make room" (Butler, 2004, p. 224)," Butler establishes a political action that challenges the social structure that serves inequality and exclusion. Butler's philosophy is not based on theoretical assumptions, rather it is woven into the fabric of culture.

When Butler proposes the theory of gender performativity, it encourages a more intersectional understanding of gender identity that extends beyond the binary frameworks. Butler also claims that the meaning of the body is understood within a discursive paradigm. Seyla Benhabib writes about Butler's idea of the construction of the body. She argues: "the already sexed body is the epistemological equivalent of the myth of the given: just as the given can be identified only within a discursive framework, so too it is the culturally available codes of gender that "sexualize" a body and that construct the directionality of that body's desire (Benhabib, 1995, p. 21). The body is not a pre-existing natural entity but can only be known through the cultural and discursive framework and is constructed through the cultural codes of gender.

In *Gender Trouble*, Butler explores the notions of gender, power, and the body as discussed in the works of Julia Kristeva, Monique Wittig, Michel Foucault, Simone de Beauvoir, Lacan, and Sigmund Freud. It is in the last chapter of Gender Trouble that she elaborates on the idea of performativity. In *Gender Trouble*, Butler seeks to show the idea that all identities are constituted through power relations and discursive practices. For Butler, gender is not an inherent or pre-given identity; rather it is a cultural performance regulated by compulsory heterosexuality. Sara Salih, a renowned scholar in Butler studies, argues in her book that “all bodies are gendered from the beginning of their social existence” (Salih, 2002, p. 62). There is a distinct differentiation between performance and performativity in Butler’s theory of performativity. In a 1994 interview, Butler gives the pivotal distinction between performance and performativity: performance presupposes a pre-existing subject, while performativity contests the very notion of the subject (Butler, 1994, p. 33). For Butler, gender cannot be considered like a theatrical performance of an individual. Butler points out that “theatrical performances can meet with political censorship and scathing criticism, gender performances in non-theatrical contexts are governed by more clearly punitive and regulatory social conventions” (Butler, 1988, p. 527). Butler employs the term ‘performativity’ to highlight the act of ‘doing’ rather than the concept of self-being, which has real effects in society and has power to create an identity. Through performativity theory, Butler rejects the notion of the ‘self,’ which leads to an act: “gender is always a doing, though not a doing by a subject who might be said to preexist the deed” (Butler, 1990, p. 25). Butler intends to establish that the construction of a subject is not a representation of the inner self; rather, it is an effect of repetitive performance. Butler argues:

> performativeness is quite crucial, for if gender attributes and acts, the various ways in which a body shows or produces its cultural signification, are performative, then there is no preexisting identity by which an act or attribute might be measured; there would be no true or false, real or distorted acts of gender, and the postulation of a true gender identity would be revealed as a regulatory fiction (Butler,1988, p. 528).

According to Butler, the concept of the ‘I’ is an illusion, and there is no pre-given identity in an individual. Butler asserts that human beings are all subjected to frameworks of power, and all subjects are produced in this matrix of power. For Butler, “There is no gender identity behind the expressions of gender; that identity is performatively constituted by the very "expressions" that are said to be its results" (Butler, 1990, p.33). Butler follows Nietzsche's philosophy of the metaphysics of substance to display that sex and gender are merely illusions of the subject. Butler needs to deny the ontological status of gender, which leads to gender performance. Here Butler follows Nietzsche's argument that "there is no 'being' behind doing, acting, becoming; 'the doer' is merely a fiction imposed on the doing — the doing itself is everything" (Nietzsche, 1887/1996, p. 29). For Butler, the performance of a performer is paramount, and the subject is constituted through this performance. The performance is an ongoing process that has no definitive end.

Through performativity, Butler was not proposing an abstract idea or theoretical model in philosophy; rather, she was more concerned with human existence as it is affected by social conditioning. Performativity does not imply that there is an actor who chooses to act according to a predefined script. Butler explains that the performance of various gender acts may differ from one performer to another.

Conclusion

This study has traced the evolving conception of the self, beginning with the Enlightenment ideals of rational agency in Mary Wollstonecraft's writings, moving through the critiques of domesticity and feminine mystique in the mid-20th century with Betty Friedan, and the nuanced exploration of gendered interiority and creative autonomy in Virginia Woolf's work. These thinkers laid the groundwork for understanding the self as a historically situated and gendered subject. Building upon this, Judith Butler's philosophy marks a significant shift by theorizing the subject not as a fixed identity but as a performative and socially constituted construct. Butler's intervention highlights that subjectivity is not pre-given but continuously produced through repeated social performances and discursive norms. The journey from the autonomous self to the postmodern subject thus reflects a deepening critique of essentialism, revealing identity as dynamic, contested, and deeply embedded in cultural practices.

References

Benhabib, S. (1995). Feminism and postmodernism. In L. Nicholson (Ed.), *Feminist contentions: A philosophical exchange* (pp. 17–34). Routledge.

Budgeon, S. (2011). *Third wave feminism and the politics of gender in late modernity*. Palgrave Macmillan.

Butler, J. (1988). Performative acts and gender constitution: An essay in phenomenology and feminist theory. *Theatre Journal, 40*(4), 519–531.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity.* Routledge.

Butler, J. (1994). Gender as performance: An interview with Judith Butler. *Radical Philosophy: A Journal of Socialist and Feminist Philosophy, 67*, 32–39.

Butler, J. (2004). *Undoing gender*. Routledge.

De Beauvoir, S. (2011). *The second sex* (H. M. Parshley, Trans.). Vintage Books. (Original work published in 1949)

Descartes, R. (1968). *Discourse on method and the meditations* (F. E. Sutcliffe, Trans.). Penguin Books. (Original work published 1637).

Friedan, B. (1963). *The feminine mystique*. Penguin Books.

Friedman, M. (2000). Feminism in ethics: Conceptions of autonomy. In M. Fricker & J. Hornsby (Eds.), *The Cambridge companion to feminism in philosophy* (pp. 216–235). Cambridge University Press.

Nietzsche, F. (1996). *On the genealogy of morals* (D. Smith, Trans.). Oxford University Press. (Original work published 1887)

Salih, S. (2002). *Judith Butler*. Routledge.

Sartre, J. P. (2003). *Being and nothingness: An essay on phenomenological ontology* (H. E. Barnes, Trans.). Routledge. (Original work published 1943)

Segal, L. (1999). *Why feminism? Gender, psychology, politics*. Polity Press.

Snyder, R. C. (2008). What is third-wave feminism? A new directions essay. *Signs: Journal of Women in Culture and Society, 34*(1), 175–196.

Stanton, E. C., Anthony, S. B., & Gage, M. J. (Eds.). (1881). *History of woman suffrage* (Vol. 1). Fowler & Wells.

Walters, M. (2005). *Feminism: A very short introduction*. Oxford University Press.

Wollstonecraft, M. (1982). *A vindication of the rights of woman*. W. W. Norton. *(Original work published 1792)*

Woolf, V. (1977). *A room of one's own*. Grafton. (Original work published 1929)

Žižek, S. (2000). Class struggle or postmodernism? Yes, please! In J. Butler, E. Laclau, & S. Žižek, *Contingency, hegemony, universality: Contemporary dialogues on the Left* (pp. 90–135). Verso.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 212-229

Ensayo/Essay

Reframing suffering: buddhist mindfulness techniques as tools for philosophical counselling

Reenmarcando el sufrimiento: técnicas budistas de atención plena como herramientas de asesoramiento filosófico

Richa Singh

Central University of Allahabad, India
1003.richasingh@gmail.com
https://orcid.org/0009-0003-3383-268X

DOI https://doi.org/10.48204/2805-1815.8483

INFORMACIÓN DEL ARTÍCULO

Recibido el: 23/07/2025

Aceptado el: 29/9/2025

Keywords:

Philosophical counselling, Buddhist mindfulness, value-based therapy, existential suffering, cognitive therapy

Palabras clave:

Asesoramiento filosófico, atención budista, terapia basada en valor, sufrimiento existencial, terapia cognitiva

ABSTRACT/RESUMEN

Abstract:

This study investigates the role of Buddhist mindfulness techniques as foundational values in the emerging field of philosophical counselling. Bridging Eastern contemplative traditions and Western philosophical practice, the paper argues that mindfulness understood not merely as meditation but as active, value-oriented awareness; can significantly contribute to the goals of philosophical counselling. Both traditions prioritize self-awareness, ethical reflection, and the alleviation of suffering through insight rather than clinical diagnosis. Drawing on the Buddhist concept of the “Second Arrow,” the paper illustrates how mindfulness can help individuals differentiate between inevitable pain and the optional suffering caused by reactive thought patterns.

The research further explores how mindfulness, grounded in the Pali concept of sati, encompasses memory, attentiveness, and ethical clarity, making it a potent tool for value-based dialogue and emotional clarity and resilience. By situating mindfulness within the framework of non-clinical, philosophical dialogue, the study challenges conventional therapeutic models and highlights a humanistic, integrative approach to counselling. This paper proposes that the synthesis of Buddhist mindfulness and philosophical counselling not only enhances individual well-being but also contributes to a broader discourse on wisdom, agency, and ethical living in contemporary society.

Resumen:

Este estudio investiga el papel de las técnicas budistas de atención plena como valores fundamentales en el campo emergente del asesoramiento filosófico. Tendiendo un puente entre las tradiciones contemplativas orientales y la práctica filosófica occidental, el artículo sostiene que la atención plena, entendida no solo como meditación sino como conciencia activa y orientada a los valores, puede contribuir significativamente a los objetivos del asesoramiento filosófico. Ambas tradiciones dan prioridad a la conciencia de uno mismo, la reflexión ética y el alivio del sufrimiento a través de la introspección, en lugar del diagnóstico clínico. Basándose en el concepto budista de la "segunda flecha", el artículo ilustra cómo la atención plena puede ayudar a las personas a diferenciar entre el dolor inevitable y el sufrimiento opcional causado por patrones de pensamiento reactivos.

La investigación explora además cómo la atención plena, basada en el concepto 'Pali de sati', abarca la memoria, la atención y la claridad ética, lo que la convierte en una herramienta potente para el diálogo basado en valores y la claridad emocional y la resiliencia. Al situar la atención plena en el marco del diálogo filosófico no clínico, el estudio desafía los modelos terapéuticos convencionales y destaca un enfoque humanista e integrador de la terapia. Este artículo propone que la síntesis de la atención plena budista y la terapia filosófica no solo mejora el bienestar individual, sino que también contribuye a un discurso más amplio sobre la sabiduría, la agencia y la vida ética en la sociedad contemporánea,

Introduction

Although counselling is often saw as a modern practice, its fundamental essence; guiding individuals through challenges, has ancient roots. Religious teachers, philosophers, and sages have long addressed human suffering through dialogue and reflection. In this context, the Buddha stands out not merely as a spiritual leader but as a Bhaiṣajyaguru - a master physician, or "Medicine Buddha", whose teachings offered psychological and existential relief. The Dhamma which he taught was like a medicine which cured the problem of suffering of the people and showed them the way leading to enduring happiness. He was concerned with the moral psychological problems of the people which assumed various forms. He found out the roots of the problems which were diverse or complex in nature and he found out their solutions also in diverse ways. On the other hand, medical practitioners have been doing counselling for curing diseases, which involves advice regarding lifestyle and application of Medicine but at the end, all therapists, including Philosophical counsellor do have similar aims.

Both Buddhist mindfulness techniques and philosophical counselling seek to alleviate suffering by fostering introspection, moral clarity, and self-awareness. These traditions emphasize dialogue, critical inquiry, and the cultivation of values distinct from medicalized or pathologized treatments. This paper argues that mindfulness-based approaches rooted in Buddhist thought offer practical, value-oriented frameworks that align seamlessly with the aims of philosophical counselling.

Philosophical Counselling: Foundations and Principles

Philosophical counselling emerged in the 1980s as a non-clinical approach to addressing existential and moral problems through reasoned dialogue and reflective inquiry. Jon Kabat-Zinn gives the most popular definition of Philosophical Counseling. In his word, "Paying attention in a particular way: on purpose, in the present moment, and non-judgmentally" (Kabat-Zinn, 1990, p. 4). Rather than diagnosing pathology, it aims to cultivate clarity, meaning, and ethical understanding.

Schuster (2008) defines philosophical counselling as a dialogical method that uses philosophical tools to help individuals reflect on life events and choices. It empowers clients to explore dilemmas not through medication or diagnosis, but through deeper questioning of beliefs, values, and assumptions.

Three core tenets underpin philosophical counselling (Li, 2010):

1. Non-pathologizing approach: Clients are not “patients” but individuals confronting philosophical challenges.

2. Leading by values: Counselling aims to clarify and guide individuals by their value systems.

3. Dialogical method: The process is conducted through structured, thoughtful dialogue rather than prescriptive treatment.

The importance of this fundamental principle lies in the emphasis that one's sufferings or problems are caused by confusion in ideas, and the clarification of ideas could help with relieving one's sufferings and problems. One of the clearest distinctions from psychotherapy lies in this commitment to philosophical dialogue. For example, consider the case of a monk experiencing depression due to a conflict between spiritual vows and family desire. While traditional psychiatric interventions failed, philosophical counselling enabled him to make meaning of the situation and transition from confusion to clarity through existential reflection. This case demonstrates that value conflicts, often mistaken for psychiatric disorders, can be resolved through philosophical inquiry.

As Martin (2006) explains, such cases often involve moral conflict rather than medical illness. Here, tools from philosophy; like Socratic dialogue, virtue ethics, or existential thought, are better suited than clinical models.

In this context Neimeyer states:

> Common sense tells us that persons suffering from symptoms of major depression-fatigue, sleeping problems, feelings of hopelessness, helplessness, and suicidal ideation, with a severity that sends them to seek professional help, are not in peak mental health. This is in accord with the authors' cognitive therapy orientation, which suggests that by changing how one thinks about or regards any event in life, we can modify the level of distress it engenders. For thanatologists, this approach also fits in well with the meaning-making or narrative approaches familiar to most practitioners in the field." (Neimeyer & Sands, 2011, pp. 9-22).

Philosopher Lou Marinoff (2001) classifies depression into four types, based on etiology: (1) genetic, (2) substance-induced, (3) trauma-based, and (4) existential, arguing that existential crises often benefit more from philosophical counselling than from medication or psychotherapy.

Philosophical counselling is particularly effective for the fourth type, when depression arises from moral or value-based dilemmas. This is especially true when the suffering stems from moral ambiguity or life transitions rather than clinical pathology. This approach overlaps significantly with mindfulness-based techniques drawn from Buddhist philosophy.

According to Marinoff, the first type of depression is a physical illness requiring the help of psychiatrists or other physicians. The second type is "a physical or psychological dependency" that also requires medical attention. The last two types of depression can benefit from "talk therapy." Specifically, the third type can benefit from psychology and sometimes from philosophical counselling. "But in the fourth scenario--by far the most common one brought to counsellors of all kinds--philosophy would be the most direct route to healing." (Martin, 2001.) In this regard, Fromm (2002) distinguishes between two types of meditative techniques that have been used in psychotherapy: (i) auto-suggestion used to induce relaxation; and (ii) meditation "to achieve a higher degree of non-attachment, of non-greed, and of non-illusion; briefly, those that serve to reach a higher level of being" (Fromm, 2002, p. 50). Fromm attributes techniques associated with the latter to Buddhist mindfulness practices. (Fromm, 2002)

Yet what distinguishes philosophical counselling is its intentional grounding in wisdom traditions, where healing is not merely about symptom reduction but about the cultivation of ethical clarity and self-understanding. This is where philosophical counselling naturally overlaps with Buddhist mindfulness, especially as interpreted in traditions that emphasize insight (*vipassanā*) over merely calming the mind (*samatha*). Both traditions encourage a reflective stance toward suffering, values, and identity. This makes it particularly compatible with Buddhist mindfulness practices, which prioritize awareness, non-reactivity, and liberation from illusion. Buddhistic mindfulness practices have been explicitly incorporated into a variety of psychological treatments. More specifically psychotherapies dealing with cognitive restructuring share core principles with ancient Buddhistic antidotes to personal suffering.

Furthermore, this emphasis on values, dialogue, and clarity also underpins Buddhist mindfulness, particularly as reframed in cognitive and therapeutic contexts. The next section explores how these traditions intersect conceptually and methodologically.

Buddhist Mindfulness: Classical Foundations and Interpretive Debate

Mindfulness, or *Sati* in Pāli (*Smṛti* in Sanskrit), is a foundational element of Buddhist thought. Often translated as "bare attention" by Nyanaponika Thera, its deeper meaning encompasses clear comprehension (*sampajañña*), vigilance (*apramāda*), and remembrance of the Dhamma (Van Gordon et al., 2014; Sharf, 2014). All three terms are sometimes (confusingly) translated as "mindfulness", but they all have specific shades of meaning. Georges Dreyfus has also expressed unease with the definition of mindfulness as "bare attention" or "nonelaborative, nonjudgmental, present-centered awareness", stressing that mindfulness in Buddhist context means also "remembering", which indicates that the function of mindfulness also includes the retention of information.

According to Bryan Levman, "the word Sati incorporates the meaning of 'memory' and 'remembrance' in much of its usage in both the suttas and the [traditional Buddhist] commentary, and ... without the memory component, the notion of mindfulness cannot be properly understood or applied, as mindfulness requires memory for its effectiveness" (Levman, 2017, p. 21). However, what does mindfulness really mean?

Bhikkhu Bodhi (2011) clarifies that while Sati originally connoted memory, the Buddha repurposed it to signify lucid awareness. This awareness is both ethical and cognitive, enabling one to remember one's values and intentions in each moment. He stated that:

> But we should not give this [meaning of memory] excessive importance. When devising a terminology that could convey the salient points and practices of his own teaching, the Buddha inevitably had to draw on the vocabulary available to him. To designate the practice that became the main pillar of his meditative system, he chose the word sati. But here sati no longer means memory. Rather, the Buddha assigned the word a new meaning consonant with his own system of psychology and meditation. Thus, it would be a fundamental mistake to insist on reading the old meaning of memory into the new context.… I believe it is this aspect of sati that provides the connection between its two primary canonical meanings: as memory and as lucid awareness of present happenings.… In the Pali suttas, sati has still other roles in relation to meditation, but these reinforce its characterization in terms of lucid awareness and vivid presentation. (Bodhi, 2011.)

In the Satipaṭṭhāna-sutta the term *Sati* means to remember the dharmas, whereby the true nature of phenomena can be seen, which means, mindfulness is a quality that every human being already possesses, it's not something you have to conjure up, you just have to learn how to access it. The Theravada Nikayas prescribe that one should establish mindfulness (*satipaṭṭhāna*) in one's day-to-day life, maintaining as much as possible a calm awareness of the four *upassanā*: one's body, feelings, mind, and dharmas, such as,

- *Kāyānupassanā* (the six sense-bases which one needs to be aware of)
- *Vedanānupassanā* (contemplation on vedanās, which arise with the contact between the senses and their objects)
- *Cittānupassanā* (the altered states of mind to which this practice leads)
- *Dhammānupassanā* (the development from the five hindrances to the seven factors of enlightenment)

The four upassanā have been misunderstood by the developing Buddhist tradition, including Theravada, to refer to four different foundations. These practices aim not only to develop concentration but also to cultivate insight into the impermanent, unsatisfactory, and non-self-nature of experience—core principles in Buddhist soteriology (Polak, 2011).

Furthermore, here I would refer to the *Milindapañha*, which explained that the arising of *Sati* calls to mind the wholesome dhammas. (Sharf, 2014) It means "moment to moment awareness of present events", but also "remembering to be aware of something". In this context Buddhadasa said, “the aim of mindfulness is to stop the arising of disturbing thoughts and emotions, which arise from sense-contact.” (Buddhadasa, 2014, p. 115) Although, according to American Buddhist monk Bhante Vimalaramsi's (2015) , the term mindfulness is often interpreted differently than what was originally formulated by the Buddha. In the context of Buddhism, he offers the following definition:

> Mindfulness means to remember to observe how mind's attention moves from one thing to another. The first part of Mindfulness is to remember to watch the mind and remember to return to your object of meditation when you have wandered off. The second part of Mindfulness is to observe how mind's attention moves from one thing to another. (Bhante Vimalaramsi, 2015, p. 4)

However, the mechanisms that make people less or more mindful have been researched less than the effects of mindfulness programs, so little is known about which components of mindfulness practice are relevant for promoting mindfulness. In order to answering these here we present the concept of philosophical counselling and the reason behind is, these both concept (counselling and Buddhists mindfulness) based on a unique subject matter and goal that aims to assist people to deal with life events in an effective manner and aimed at wisdom.

Buddhist Counselling as Value-Oriented Philosophy

Buddhism, particularly in the Mahayana tradition, introduces the Bodhisattva ideal, adding a collective, compassionate dimension to counselling. Here, personal healing is interwoven with ethical responsibility and altruism. Mindfulness, in this context, becomes a means of awakening, not just a stress-reduction tool.

Contemporary interpretations, especially in the West, have often reduced mindfulness to a stress-reduction technique or emotional regulation tool. While these applications are valuable, they risk neglecting the full ethical and philosophical dimensions of sati. Teachers like Jon Kabat-Zinn have sought to reintroduce mindfulness into secular contexts (e.g., in Mindfulness-Based Stress Reduction), but even Kabat-Zinn acknowledges its Buddhist roots and universal aims of reducing suffering and dispelling delusion (Kabat-Zinn, 2005).

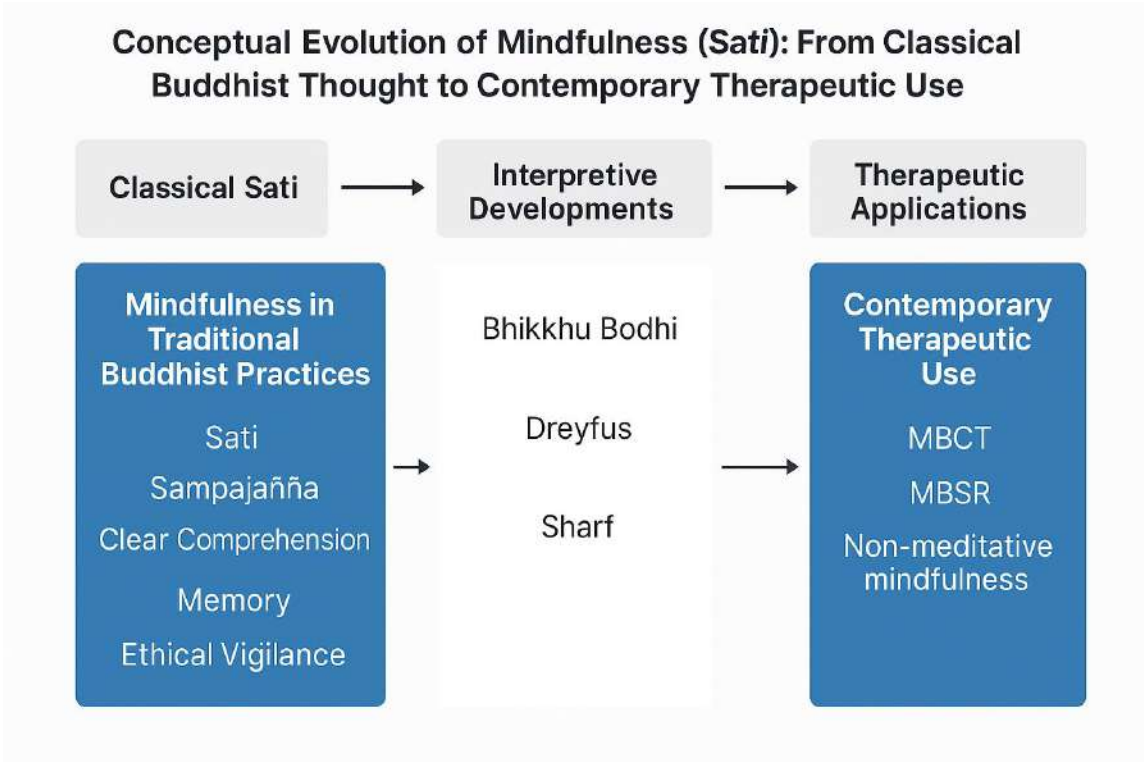
Other scholars, such as Georges Dreyfus and Robert Sharf, have criticized the oversimplified definition of mindfulness as “non-judgmental present awareness,” arguing that such terms overlook its ethical backbone. In Buddhist practice, mindfulness is inseparable from intention, right view, and the pursuit of awakening.

The mindful pause can prevent misperceptions from arising. As Ruth King noted: "Simply stated, we perceive something through our senses. There is a sense organ, and a sense object-eyes see, ears hear, nose smells, body feels, tongue tastes, and mind thinks. Once we perceive, we habitually jump to thoughts and feelings about what is being perceived. These thoughts and feelings, rooted in past experiences and conditioning, then influence the mood of our mind. When perception, thoughts, and feelings are repeated or imprinted through experiences, they solidify into view or belief. View then reinforces perception. This cycle becomes the way we experience and respond to the world." (King, 2018)

Moreover, mindfulness is not limited to meditation. It is a life skill applicable to speech, action, and daily conduct, supporting the philosophical counselling ideal of living an examined, ethical life. Thus, philosophical counselling can benefit immensely from Buddhist insights, offering pluralistic, non-pathologizing frameworks that promote growth, clarity, and inner freedom.

Thus, any integration of mindfulness into philosophical counselling must recover these deeper dimensions. Mindfulness is not only a method for calming the mind—it is a lens through which one discerns reality, questions attachments, and makes value-aligned choices; this orientation that makes mindfulness a powerful complement to philosophical dialogue.

Figure: Conceptual Evolution of Mindfulness (Sati)



Mindfulness Beyond Meditation: Talk-Based Techniques

While contemporary interpretations of mindfulness often focus on meditation, traditional Buddhist contexts also emphasize non-meditative applications. Techniques such as mindful speech, ethical behaviour, and introspective dialogue are pivotal in therapeutic and philosophical interventions.

The closest words for meditation in the classical languages of Buddhism are *bhāvanā* ("mental development"). *Bhāvanā* can involve cultivating virtues such as patience, forbearance, equanimity, wisdom, and compassion. *Vipassanā* and *samatha* are described as qualities which contribute to the development of mind (*bhāvanā*). *Vipassanā* is commonly used as one of two poles for the categorization of types of Buddhist practice, the other being *samatha*. Various traditions disagree which techniques belong to which pole. (Schumann, 1974.) According to the contemporary *Theravada* orthodoxy, *samatha* is used as a preparation for *vipassanā*, pacifying the mind and strengthening the concentration in order to allow the work of insight, which leads to liberation. Though both terms appear in the Sutta Pitaka, Gombrich argues that the distinction as two separate paths originates in the earliest interpretations of the Sutta Pitaka, not in the suttas themselves. (Gombrich, 1997)

In Dialectical Behavior Therapy (DBT) and Acceptance and Commitment Therapy (ACT), non-meditative mindfulness exercises help clients identify and reframe their thoughts and emotions (Linehan, 1993). These approaches mirror Buddhist practices like Right Mindfulness (*Sammā-Sati*) and Right View (*Sammā-Diṭṭhi*), both crucial in navigating suffering without aversion and/or attachment.

Such methods align closely with philosophical counselling's emphasis on value clarification, self-awareness, and narrative reconstruction—tools that enable clients to reframe distressing experiences as opportunities for growth and ethical reflection.

Furthermore, when we go from Theravāda Buddhism to Mahāyāna, a new dimension is added to the Buddhist counselling. It is that of Bodhisattva's altruistic mission. Here I want to consider Mahayana way of counselling as a part of diversity and not a part of hierarchy. It is based on the idea that suffering is a part of life, and that problems can be worked through to achieve freedom from suffering. A Buddhist counselling approach that uses mindfulness and cognitive training to help clients understand stressful situations. Non-meditation-based mindfulness exercises are specifically used in dialectical behaviour therapy. It may also incorporate traditional talk therapies like Cognitive Behavioural Therapy (CBT) and Acceptance and Commitment Therapy (ACT). (Linehan, 1993)

The use of mindfulness-based cognitive therapy and mindfulness-based stress reduction is supported by the strongest level of evidence. To show this, we can take here a Buddhist parable from Sallatha Sutta, called- 'The Second Arrow of Buddha' as an example. According to the core Buddhist psychology models of the "Two Arrows of Pain" and "Co-dependent Origination" (PRATITYASAMUTPADA), pain is the resultant of bodily and mental factors, which can be regulated by meditation states and traits. Here we investigated how pain and the related aversion and identification (self-involvement) experiences are modulated by focused attention meditation (FAM), open monitoring meditation (OMM), and loving kindness meditation (LKM), as well as by meditation expertise.

The Parable of the Second Arrow: A Philosophical Analogy

A pivotal Buddhist teaching—the Parable of the Second Arrow—illustrates the difference between pain and suffering (Sallatha Sutta, SN 36.6). The Buddha explains: In Buddhist teachings, the parable of the second arrow goes as follows:

The Buddha once asked a student, "If a person is struck by an arrow, is it painful?

The student replied, "It is" The Buddha then asked, "If the person is struck by a second arrow, is that even more painful? " The student replied again, "It is". The Buddha then explained, "In life, we cannot always control the first arrow. However, the second arrow is our reaction to the first. And with this second arrow comes the possibility of choice. The Buddhists say that any time we suffer misfortune; two arrows fly our way. Being struck by an arrow is painful. Being struck by a second arrow is even more painful.

This second arrow symbolizes the mental elaboration—self-blame, resistance, fear—that transforms inevitable pain into avoidable suffering. As Haruki writes: "Pain is inevitable; suffering is optional." (Haruki, 2009, p. 7) This pain turns into suffering in its extreme stage. We get to see a clear distinction between pain and suffering under this parable of second arrow of Buddhism. Well, it said that according to modern psychology (not to mention ancient Buddhism), therein lies the difference between pain and suffering. However, the two are not the same thing! Pain is what happens to us, suffering is what we do with that pain. While changing our perception of this concept may be difficult, it is possible. We can avoid or lessen our actual suffering based on what we choose to do with the pain we experience.

Buddhism teaches that the fundamental source of all suffering is this very attachment to or aversion to experience (Bercholz & Kohn, 1993.). For example, if we lose a loved one, we cannot get rid of that pain, but instead of asking ourselves, why did this happen to me? Could I have saved them? Instead, we can say to ourselves that I am not the only one with whom this has happened, I have done what I could, and I should try to do my best in such situation in future also.

There is a sense of resistance to it - not accepting it, not allowing it to be there, and accepting the reality of the situation. We fight with the reality of the way things are right now and so we turn the pain into suffering, or we add suffering on top. There is an equation that is often used in ACT which is:

Pain x Resistance = Suffering

The more that you resist or deny or fight or argue with the pain- which is already there, the more suffering you experience. That is a useful story to remember whenever you have any kind of demanding situation. It could be difficult internal experiences - there could be difficult emotions like sadness, anxiety, frustration, or anger, or it could be to do with difficult thoughts; it could be difficult sensations like literal physical pain or chronic pain.

Although, we are only human, and we may be overcome by feelings before we know it, even that is life. It would be abnormal not to experience feelings when major events happen. A mindfulness training will get us through the punctured tire unscathed, but major events in life like birth, death, disease, or divorce will not be always overcome with merely a meditative attitude. These kinds of arrows will also hit us, eventually. It will not always be possible to prevent the second arrow from hitting us. We will be sad, angry, afraid, even have self-pity, be depressed, etc. As Buddhist nun, Pema Chodron has suggested "Meditation practice isn't about trying to throw ourselves away and become something better. It is about befriending who we are already. The ground of practice is you or me or whoever we are right now……that's what we study, that's what we come to know with tremendous curiosity and interest." (Chodron, 1993, p. 27)

Thus, mindfulness enables a pause between stimulus and reaction. Viktor Frankl (2017) articulated this gap as the essence of human freedom: "Between stimulus and response, there is a space. In that space is our power to choose our response."

Our power is in the space that we can create between stimulus and response. Creating that space is the key to avoiding the second arrow. Here we find another equation that is we can use as a treatment, which is:

Pain x Acceptance = Freedom

As Frankl famously said, "Everything can be taken from a man but one thing: the last of the human freedoms—to choose one's attitude in any given set of circumstances, to choose one's own way." (Frankl, 2017, p. 55)

Let us get an example, we cannot even imagine the pain that a woman feels while giving birth to a child, but still that labour pain is not become suffering for her because her pleasant experience is associated with that pain. At that time, above any negative situation in his mind, there is a feeling of happiness that he gets from seeing his child. It never means that his pain is less in any respect, but his positive thoughts related to that pain stop that pain from turning into suffering. In other circumstances, it may not be possible that we can associate any positive thought with pain, but we can try to stop that negative thought associated with pain, so that the pain remains but does not turn into suffering. And while we cannot control our outside environment, we can, with practice, change this pattern of shooting a second arrow after the first. There are two highly effective exercises which can practice in order to circumvent this all-too-human response to life. First, noticing the pattern of the second arrow; second, practicing kindness to yourself when you see it.

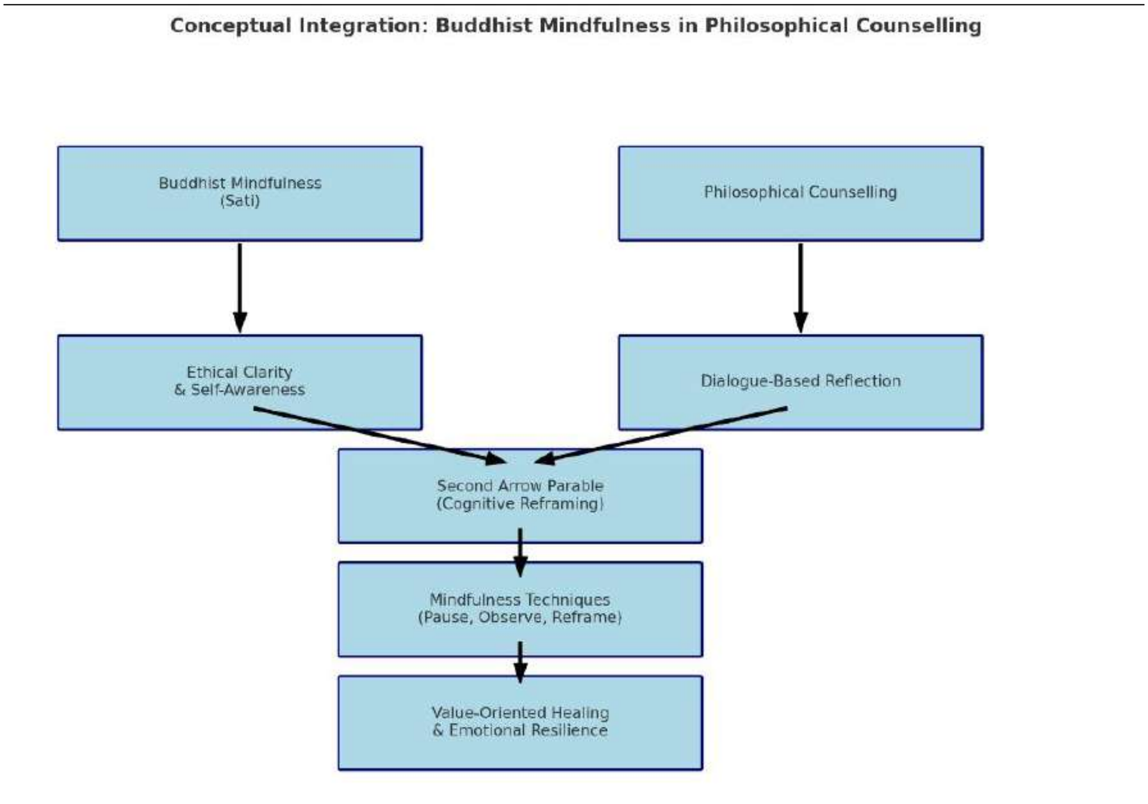
Integrating Mindfulness into Philosophical Counselling

The integration of Buddhist mindfulness into philosophical counselling is not merely additive—it is synergistic. Both traditions are oriented toward self-awareness, value-based reflection, and freedom through insight, rather than symptom eradication. Key areas of alignment include:

- **(i) Ethical Grounding:** Mindfulness in Buddhism is inseparable from ethics (sīla). Similarly, philosophical counselling helps clients examine and act in alignment with their values.
- **(ii) Dialogue over Diagnosis:** Like Buddhist teacher-student inquiry (kathā), philosophical counselling privileges open-ended dialogue and questioning over labels.
- **(iii) Transforming Suffering:** Both traditions view suffering not as pathology but as an opportunity for growth and wisdom.
- **(iv) Agency and Freedom:** Mindfulness emphasizes awareness and choice; counselling emphasizes autonomy and responsibility.

Importantly, mindfulness here is not limited to formal meditation. As Buddhadasa and others suggest, it includes mindful speech, ethical action, and attentiveness in daily life. This “talk-based mindfulness” aligns well with counselling conversations and can be directly employed in sessions without requiring meditative training.

Furthermore, Mahāyāna Buddhism adds the Bodhisattva ethic, a commitment to collective well-being. This dimension expands the scope of counselling from personal relief to social compassion, aligning with humanistic and dialogical values in contemporary philosophy.



A Value-Based Therapeutic Dialogue: Practical Techniques

Mindfulness-based philosophical counselling does not require esoteric practices. It offers a grounded set of conversational tools rooted in self-reflection, presence, and ethical clarity. Instead of demanding advanced meditative training, here some simple techniques can be adapted into counselling practice:

1. **Pause and observe:** Encourage clients to close their eyes and tune into inner dialogue. Ask them to momentarily stop, breathe, and tune into their inner experience. Even a brief pause can interrupt reactive patterns.
2. **Differentiate arrows:** Help clients distinguish between identify the primary pain (fact) and secondary suffering (reaction). Primary pain: The objective fact (e.g., loss, failure). Secondary suffering: The emotional reactivity or judgment layered upon the fact. Ask: "What is the fact, and what is my story about it?"

3. **Reflect and reframe:** Guide clients to examine habitual thoughts, like ask; “What am I telling myself about this?” & “Is this narrative helpful or harmful?” This reflects the Buddhist principle of *Yoniso Manasikāra* (wise attention).
4. **Cultivate acceptance:** Emphasize self-compassion, non-judgment, and presence. Introduce the formula: Pain × Acceptance = Freedom Acceptance is not resignation, but an ethical stance of compassion and clarity. Clients learn to respond rather than react.
5. **Respond mindfully:** Shift from reactivity to purposeful, value-aligned action to create Self-Kindness as Ethical Grounding. Encourage clients to relate to themselves as they would to a friend. As Phödrön (1993) writes: “Meditation practice is not about throwing ourselves away and becoming something better. It is about befriending who we are.”

These steps offer clients tools to deconstruct habitual suffering patterns and access clarity, presence, and ethical orientation, hallmarks of both mindfulness and philosophical living. These techniques illustrate how mindfulness can be applied in counselling without formal meditation. When clients begin to observe their thoughts and emotions with non-judgmental awareness, they start to shift from identification to insight; from suffering to understanding.

Not surprisingly, in terms of clinical diagnoses, MBSR has proven beneficial for people with depression and anxiety disorders; however, the program is meant to serve anyone experiencing significant stress. (Segal et al., 2002.) Although, this is just an example. There are various of mindfulness-based teaching available in Buddhists tradition, which can truly lead the significant role in the perspective of philosophical counselling. Kabat-Zinn, a one-time Zen practitioner, goes on to write: "Although at this time, mindfulness meditation is most commonly taught and practiced within the context of Buddhism, its essence is universal. Yet it is no accident that mindfulness comes out of Buddhism, which has as its overriding concerns the relief of suffering and the dispelling of illusions". (Kabat, 2009, pp. 12-13)

Conclusion

This study has explored how Buddhist mindfulness; particularly when understood in its original context as sati, or ethical awareness, can be meaningfully integrated into the practice of philosophical counselling. The dialogue between these two traditions reveals deep conceptual harmony: both reject pathologizing models, emphasize the cultivation of self-awareness, and focus on values as a means to address suffering.

The parable of the Second Arrow provided a central framework for this dialogue. By helping individuals distinguish between unavoidable pain and avoidable suffering, it offers a cognitive and ethical tool that empowers clients to recognize and deconstruct reactive patterns. In philosophical counselling, this becomes not only a technique but a principle: to shift from habitual reaction to thoughtful reflection grounded in personal meaning.

The paper also highlighted how non-meditative mindfulness techniques, such as mindful speech, inquiry-based reflection, and value clarification, can enhance therapeutic conversations. These methods are accessible, culturally adaptive, and consistent with both Buddhist and philosophical aims: liberation from confusion, living ethically, and acting with clarity.

Moving forward, more empirical research is needed to explore how these integrated practices affect client outcomes. Comparative studies between mindfulness-based cognitive therapy (MBCT) and mindfulness-informed philosophical counselling could offer valuable insights. Additionally, developing training modules that blend these traditions may help create more inclusive and reflective counselling practices.

Ultimately, integrating Buddhist mindfulness into philosophical counselling is not just about borrowing techniques—it is about rethinking counselling itself as a wisdom-based practice, where healing emerges not through control or correction but through awareness, acceptance, and ethical dialogue.

References

Bercholz, S., & Kohn, S. C. (1993). *An introduction to the Buddha and his teachings*. Barnes & Noble.

Bhante Vimalaramsi. (2015). *A guide to tranquil wisdom insight meditation*. CreateSpace Independent Publishing.

Bodhi, B. (2011). What does mindfulness really mean? A canonical perspective. In *Investigating the Dhamma*.

Buddhadasa Bhikkhu. (2014). *Heartwood of the Bodhi Tree*. Wisdom Publications.

Chödrön, P. (1993). *The wisdom of no escape and the path of loving-kindness*. Shambhala.

Frankl, V. E. (2017). *Man's search for meaning*. Beacon Press.

Fromm, E. (2002). *The art of being*. Continuum.

Gombrich, R. F. (1997). *How Buddhism began: The conditioned genesis of the early teachings*. Munshiram Manoharlal.

Haruki, M. (2009). *What I talk about when I talk about running*. Vintage Book edition.

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. Dell Publishing.

Kabat-Zinn, J. (2005). *Wherever you go, there you are: Mindfulness meditation in everyday life.* (10th anniversary edition) Hatchette Books.

King, R. (2018). *Mindful of race: Transforming racism from the inside out*. Sounds True.

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Press.

Levman, B. (2017). Putting smṛti back into sati: Putting remembrance back into mindfulness. *Journal of the Oxford Centre for Buddhist Studies*, 13, 121–150.

Li, B. (2010). The three fundamental principles of philosophical counseling.

Martin, M. W. (2001). Ethics as therapy: Philosophical counseling and psychological health. *International Journal of Philosophical Practice*, 1(1).

Martin, M. W. (2006). From morality to mental health: Virtue and vice in therapeutic culture. Oxford University Press.

Neimeyer, R. A., & Sands, D. C. (2011). Meaning reconstruction in bereavement: From principles to practice. In R. A. Neimeyer et al. (Eds.), *Grief and bereavement in contemporary society* (pp. 9–22). Routledge.

Polak, G. (2011). *Reexamining jhana: Towards a critical reconstruction of early Buddhist soteriology*. UMCS.

Schuster, S. (2008). Philosophical counselling. *Journal of Applied Philosophy*, 8, 219–223.

Schumann, H. W. (1974). *Buddhism: An outline of its teachings and schools*. Theosophical Publishing House.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. Guilford Press.

Sharf, R. (2014). Mindfulness and mindlessness in early Chan. *Philosophy East and West*, 64(4), 933–964.

Van Gordon, W., Shonin, E., Griffiths, M. D., & Singh, N. N. (2014). There is only one mindfulness: Why science and Buddhism need to work together. *Mindfulness*, 6, 49–56. https://doi.org/10.1007/s12671-014-0379-y

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 230-248

Ensayo/Essay

Williams and Dussel on opacity: toward a non-totalizing method

Williams y Dussel sobre la opacidad: hacia un método no totalizador

Chris Sawyer

Independent Researcher, New York, United States
cs154@caa.columbia.edu
https://orcid.org/0009-0006-2651-9398

DOI https://doi.org/10.48204/2805-1815.8486

INFORMACIÓN DEL ARTÍCULO

Recibido el:16/7/2025

Aceptado el: 29/9/2025

Keywords:

Opacity, methodology, ethical theory, totality, structuralism

Palabras clave:

Opacidad, metodología, teoría ética, totalidad, estructuralismo

ABSTRACT/RESUMEN

Abstract:

This paper examines the role of opacity in the thought of Bernard Williams and Enrique Dussel, and develops a philosophical method oriented toward non-totalization. In Williams's ethics, opacity marks the internal limits of moral justification: the individual cannot and should not make all aspects of their ethical life transparent to others or to themselves. Integrity, for Williams, resides not in public coherence but in the lived coherence of one's commitments, which remain partially inarticulable. In contrast, Dussel identifies opacity at the structural level—as the condition of exteriority that totalizing systems must exclude in order to sustain their coherence. Through his concept of analectics, Dussel maintains that this exclusion is not accidental but constitutive: the Other is not simply marginalized but rendered epistemically invisible. By placing these two accounts into dialogue, the paper argues that opacity should be treated not as a failure of knowledge or clarity, but as a methodological principle. This principle finds further support in recursive systems, where self-reference generates non-coincidence from within. The resulting framework affirms opacity as a condition of ethical, structural, and conceptual integrity. Rather than seeking philosophical totality, the method outlined here sustains the limits of systems as an active site of reflection. Opacity, in this sense, becomes not what philosophy must overcome, but with what it must think.

Resumen:

Este artículo examina el papel de la opacidad en el pensamiento de Bernard Williams y Enrique Dussel, y desarrolla un método filosófico orientado hacia la no-totalización. En la ética de Williams, la opacidad marca los límites internos de la justificación moral: el individuo no puede ni debe hacer completamente transparentes todos los aspectos de su vida ética, ni para los demás ni para sí mismo. La integridad, según Williams, no reside en la coherencia pública, sino en la coherencia vivida de los propios compromisos, que permanecen en parte inarticulables. En contraste, Dussel identifica la opacidad en un nivel estructural, como la condición de exterioridad que los sistemas totalizantes deben excluir para sostener su coherencia. A través de su concepto de analéctica, Dussel sostiene que esta exclusión no es accidental sino constitutiva: el Otro no es simplemente marginado, sino hecho epistemológicamente invisible. Al poner en diálogo estas dos concepciones, el artículo argumenta que la opacidad debe entenderse no como un fallo del conocimiento o de la claridad, sino como un principio metodológico. Este principio recibe apoyo adicional en sistemas recursivos, donde la autorreferencia genera no coincidencia desde dentro. El marco resultante afirma la opacidad como una condición de integridad ética, estructural y conceptual. En lugar de buscar la totalidad filosófica, el método aquí propuesto sostiene los límites de los sistemas como un sitio activo de reflexión. La opacidad, en este sentido, no es lo que la filosofía debe superar, sino aquello con lo que debe pensar.

Introduction – The Problem of Totalization

Philosophy has long been marked by the aspiration toward clarity, coherence, and systematic completeness. From the drive toward unifying metaphysical principles to the construction of normative frameworks that claim universal applicability, many philosophical traditions have assumed that to think well is to render things transparent: to eliminate ambiguity, resolve contradiction, and account for every relevant element within a single intelligible order. This aspiration—whether epistemological, ethical, or ontological—often presumes that the goal of thought is totalization. That is, philosophical systems tend to define themselves not only by what they include, but by their refusal to acknowledge what cannot be fully integrated. In this process, difference becomes deviation, ambiguity becomes error, and opacity becomes failure (Dussel, 1996; Williams, 1985).

This paper challenges that assumption. It argues that opacity is not a limit to be overcome, but a structural condition of thought itself. To think philosophically with integrity requires the ability to sustain, rather than eliminate, what cannot be totalized. Opacity here refers not to confusion or obscurity, but to a principled refusal of complete visibility—an acknowledgment that certain aspects of selfhood, relation, or system remain partially inaccessible, not because they are unexamined, but because their very structure resists total articulation. In place of a philosophy that aims for closure, this paper proposes a method that begins from structural incompletion.

This argument is developed through a comparative reading of two figures who, in distinct but complementary ways, resist the demand for totalization: Bernard Williams and Enrique Dussel. Although they emerge from different traditions—Williams from post-analytic moral philosophy and Dussel from Latin American philosophy of liberation—both thinkers articulate conceptions of ethical and structural life in which opacity plays a significant role. For Williams, opacity arises within the moral agent: it marks the limits of self-justification and the irreducibility of personal integrity (Williams, 1981). For Dussel, opacity appears at the edge of systems: it defines the relation between a totalized philosophical or political order and the excluded exteriority that makes it possible. Through his concept of analectics, Dussel maintains that this exclusion is not accidental but constitutive: the Other is not simply marginalized, but rendered epistemically invisible (Dussel, 1988).

By bringing Williams and Dussel into dialogue, the paper identifies a shared philosophical commitment: both thinkers refuse to reduce the complexity of lived experience to a fully intelligible system. They acknowledge that what remains partially inaccessible is not always a problem to be solved, but often a feature to be preserved. Williams defends the moral legitimacy of decisions and commitments that cannot be justified in terms of public reason, while Dussel insists that totalizing systems are blind to the conditions of their own construction. In both cases, opacity functions not as a regrettable deficiency, but as a condition of moral and political responsibility.

The paper then extends this insight into a third register: recursion. Drawing from systems theory and philosophical models of self-reference, it argues that opacity is not only a limit that appears at the edge of ethical or political life, but a product of internal structural dynamics. Recursive systems—systems that refer back to themselves—are inherently incomplete. They generate misalignment, delay, and non-coincidence not by accident, but by virtue of their very mode of operation (Maturana & Varela, 1980; Metzinger, 2003). The subject, understood as a recursive structure, cannot fully coincide with itself; its self-understanding is always mediated by temporal, interpretive, and structural asymmetry. In this context, opacity emerges not from ignorance or exclusion, but from the reflexive structure of identity itself.

Taken together, these three accounts—Williams's ethical opacity, Dussel's structural exteriority, and recursive non-coincidence—form the basis for a general philosophical method. Rather than treating opacity as a deficiency to be eliminated, this method treats it as a positive condition for reflection. It resists the impulse to finalize, resolve, or enclose. It accepts that systems, selves, and ideas have limits that are constitutive rather than contingent. In doing so, it reframes opacity as a philosophical resource: a way of preserving fidelity to complexity, sustaining openness to the other, and thinking from within structural incompletion.

The structure of the paper follows this trajectory. Next section examines Williams's critique of moral transparency and the ethical role of opacity in maintaining personal integrity. Third section turns to Dussel's critique of totalized philosophical systems and his proposal of analectics as a method of engaging with what lies outside their boundaries. Fourth section explores the internal dynamics of recursive systems, where self-reference produces structural non-coincidence and opacity arises from within. Fifth section synthesizes these accounts and proposes opacity as a general philosophical method, suitable for contexts in which transparency becomes an instrument of reduction or control. Sixth section concludes by reflecting on the broader implications of this method for philosophical practice.

What follows, then, is not a defense of obscurity, nor an argument against clarity. It is, rather, an effort to clarify the kinds of limits that clarity itself must respect. Opacity, when acknowledged as structural rather than accidental, opens a space for philosophical integrity—one in which the refusal of totalization becomes not a failure of rigor, but a condition of responsibility.

Opacity in Ethical Life: Bernard Williams and the Limits of Disclosure

Philosophy has long harboured the dream of complete intelligibility. Nowhere is this more evident than in moral theory, where the aspiration to render ethical life fully transparent – subject to articulation, justification, and public reason– has structured debates from Plato to contemporary moral constructivism. Yet this aspiration is not neutral. It carries with it an implicit conception of the moral agent as one who can, and should, make all things clear: to others, to institutions, to themselves. Against this vision of moral lucidity, Bernard Williams introduces an alternative sensibility—one in which ethical life is pervaded by a kind of necessary opacity.

Williams's critique of moral theory hinges on the idea that certain features of ethical life cannot be fully captured by theoretical abstraction without distortion. In *Ethics and the Limits of Philosophy*, he repeatedly challenges the pretension of moral philosophy to offer a comprehensive account of what one ought to do (Williams, 1985). For Williams, the demand for full moral transparency—particularly the demand to justify all actions in universally acceptable terms—violates the integrity of lived experience. Integrity, as he famously articulates it, is not reducible to rational coherence or impartial justification. Rather, it emerges from the coherence of one's character, projects, and self-understanding over time (Williams, 1981). The individual's relation to themselves and their history cannot always be made legible to others, nor should it be.

This emphasis on integrity reveals the ethical significance of opacity. Consider Williams's treatment of *moral luck*: the idea that moral responsibility is affected by factors outside the agent's control. In confronting this problem, Williams does not attempt to explain away luck in favour of a purified moral core. Instead, he embraces it as a constitutive feature of ethical life (Williams, 1981). To be a moral agent, for Williams, is to inhabit a world in which one's intentions, actions, and outcomes are not always aligned—and where the judgments of others may remain partial, if not opaque. The moral landscape is not a space of perfect visibility; it is structured by asymmetries, partial perspectives, and irreducible ambiguity.

Opacity here is not simply the limit of someone else's understanding. It is an internal feature of ethical life itself. The moral agent does not have complete access to their own motives, nor are they fully transparent to themselves in moments of decision. To demand full disclosure—to oneself or others—is to deny this fact and to install in its place a moral fiction. Williams's resistance to this fiction is not grounded in anti-rationalism or relativism. Rather, it is a defense of realism: a realism about the complexity of ethical life and the conditions under which human beings make decisions (Williams, 1985).

This realism resists the totalizing impulse of moral theory in two keyways. First, it affirms the irreducibility of context. Actions cannot be understood apart from the temporal, emotional, and social entanglements in which they occur. A universalizing moral theory that strips away these entanglements in the name of clarity only succeeds in abstracting from the very life it aims to evaluate. Second, it affirms the value of non-justification. Some actions, Williams suggests, are not the result of general principles but arise from a deep fidelity to one's commitments or character. To force such actions into a justificatory framework—especially one aimed at impartial spectators—is to betray the very structure of the moral self (Williams, 1981).

In both respects, opacity functions as a moral principle. It marks the boundary of what should be made visible, rather than merely the failure to render something visible. In this way, opacity is not a problem to be solved, but a structural feature to be respected. Williams does not deny that ethical discussion, justification, and reflection are important. What he denies is that they are exhaustive. Ethical life, for him, always exceeds the reach of the theories that attempt to account for it (Williams, 1985).

This insight invites a deeper philosophical reconceptualization of opacity. Rather than viewing it as a contingent or regrettable limit—a byproduct of ignorance, irrationality, or social constraint—we can begin to see opacity as constitutive. That is, we can view it as a necessary condition of ethical agency. Without opacity, the agent would be reduced to a node of transparent decision-rules or public reason procedures. The space for moral development, ambiguity, and self-formation would collapse.

The idea that opacity is constitutive of ethical life has implications beyond the individual. It bears on our understanding of social institutions, interpersonal relations, and the very form of moral discourse. In institutions, the demand for transparency often disguises a deeper impulse to normalize, discipline, or foreclose ambiguity. The bureaucratic imperative to document and justify every action undercuts the possibility of moral discretion or depth. Williams's thought, while not explicitly institutional, points toward a critique of such moral managerialism. If integrity requires space for opacity, then any system that demands total legibility will risk eroding the conditions for moral agency itself.

Interpersonally, opacity enables respect. To recognize that the other is not fully accessible—that their inner life exceeds our grasp—is not a failure of empathy but a mark of ethical maturity. The insistence on full mutual transparency, often valorized in liberal moral theory, can border on a form of epistemic aggression: a refusal to let the other remain partially veiled. Williams's emphasis on the partiality of moral understanding suggests an alternative ethos—one in which the acknowledgment of opacity becomes a condition for ethical relation, not its failure.

Importantly, opacity does not entail silence or retreat. It does not recommend a withdrawal from ethical reflection or conversation. Instead, it reconfigures the terms of such reflection. It asks us to accept that some aspects of the moral life are not fully articulable, and that this inarticulability is not a lack to be filled but a presence to be acknowledged. In this light, opacity becomes not just a feature of ethical life, but a method of philosophical thought—a mode of engaging with moral questions that resists the temptation to reduce them to transparent logic or universal code.

This methodological shift opens the way for a broader application of opacity as a structural principle. In the next sections, I will turn to Enrique Dussel's critique of totalizing reason and his proposal of *analectics*, where the exterior—the excluded, the unrepresented—forms the necessary boundary of philosophical systems. Through this, we will see how opacity can function not only within ethical self-understanding but as a mode of resisting systemic closure. The path from Williams to Dussel is not linear, but the conceptual bridge they share—an attentiveness to what escapes systematization—lays the groundwork for a general method of non-totalizing thought.

Exteriority and Analectics: Enrique Dussel's Critique of Totalized Reason

If Bernard Williams challenges the moral totalization of the individual—refusing the fantasy of full ethical transparency—Enrique Dussel poses an analogous challenge at the level of systems. Where Williams locates opacity within the self, Dussel finds it beyond the self, in the exterior: the irreducible outside that philosophical systems, institutions, and discourses exclude in order to stabilize themselves. In both cases, opacity is not merely a residue of ignorance but a structural remainder, a marker of what cannot be internalized without distortion. Dussel names this the problem of totality, and his proposed response—analectics—offers a method for encountering what lies beyond it.

Dussel's critique of totality draws on and departs from the dialectical tradition. Following Levinas and Marx, he argues that philosophical systems construct their coherence by enclosing meaning within internal operations of mediation, synthesis, or negation. The dialectic, in its classical and Hegelian forms, absorbs contradiction into ever-higher unity, leaving nothing truly outside. But for Dussel, the exterior is not reducible to an internal moment of dialectical development. It is not a lack to be resolved but a presence that escapes and conditions the totality from without (Dussel, 1988). As Rodríguez Reyes (2022) notes, this departure reflects a broader transmodern project that refuses to collapse alterity into the logic of historical synthesis. The system is not all there is. Beyond it, there remains the Other—concrete, historical, and irreducibly opaque.

This philosophical move has deep ethical and political implications. It allows Dussel to locate violence not only in acts but in the structure of philosophical thought itself. When reason claims to speak universally, to represent all perspectives from within its own system, it erases the conditions of its own constitution. Colonial reason, Eurocentric modernity, and technocratic universality become instances not just of exclusion but of self-blindness: forms of thought that mistake their own horizon for the whole. In this sense, totality is a kind of epistemic closure—not because it cannot know everything, but because it cannot see what it excludes as constitutively outside (Dussel, 1988).

The concept of exteriority serves to mark this boundary. Unlike marginality, which implies a position within the system albeit at its edges, exteriority denotes what is not assimilable: what resists incorporation without remainder. This exteriority is not an unknowable mysticism; it is structured, situated, and historically real. It is the presence of those who have been rendered invisible by the totalizing system: the colonized, the exploited, the forgotten. Crucially, the exterior is not merely outside the content of philosophy—it marks a structural opacity in the form of philosophy itself. To the extent that philosophical systems depend on internal closure, they render themselves incapable of accounting for the conditions of their own possibility.

It is here that Dussel introduces analectics as a methodological alternative. If dialectics presumes that contradiction can always be subsumed, analectics maintains that some limits are not dialectically recoverable. Instead of overcoming the Other, analectics listens to it. The exterior is not an error to be corrected but a voice to be heard. Dussel describes analectical reason as a second ethics—one that arises not from within the

ISSN – L 2805 – 1815

system but from the face of the excluded, invoking Levinas's notion of ethical transcendence, but anchoring it in material and historical structures (Dussel, 1988).

What makes Dussel's view particularly relevant to our inquiry is his insistence that the excluded Other remains unrepresentable within the system. Attempts to make the Other fully legible—to translate their position into the terms of the dominant order—risk repeating the very erasure they seek to repair. In this sense, Dussel's notion of exteriority is an epistemological opacity that serves a critical function. It refuses the demand for total intelligibility. It maintains that some dimensions of alterity cannot and should not be domesticated by philosophical clarity. To think otherwise is to reinscribe domination as method.

This methodological opacity challenges prevailing norms of philosophical justification. Where analytic traditions emphasize internal coherence and evidentiary sufficiency, Dussel insists on the importance of structural positionality—of where thought begins. For him, the validity of a claim is not separable from its situatedness: whether it emerges from the interior of the system or from the opacity of the exterior. Thought that originates from the latter cannot be measured by the standards of the former without distortion. This is not a call for relativism but for what might be termed epistemic asymmetry—a recognition that positions do not stand on equal grounds and that the transparency of one standpoint often depends on the opacity of another. Recent work by Sánchez-Pérez (2023) highlights how philosophical systems grounded in Western universality tend to misrecognize epistemic asymmetry as deficiency rather than structural location.

Opacity here is not an obstacle to philosophy but a resource for its reconstitution. Dussel does not propose to abandon systematicity or coherence altogether. Rather, he asks that we interrupt our systems at the point where they render invisible the voices that trouble them. In this sense, opacity functions as a critical threshold—not of knowledge, but of system-legibility. The Other's opacity signals the limit of systematization, the moment when the effort to understand must give way to the willingness to be addressed.

This move has clear affinities with Williams's ethics of non-totalization. Just as Williams resists the demand for full self-disclosure in the moral sphere, Dussel resists the philosophical demand to render all alterity transparent. In both cases, opacity protects something essential: the integrity of the moral agent for Williams, and the dignity of the excluded subject for Dussel. But where Williams emphasizes personal coherence, Dussel emphasizes structural relation. His philosophy is not centered on the individual but on the historically mediated asymmetry between the speaking subject and those denied voice. Thus, Dussel extends opacity from ethical agency to ontological structure.

Importantly, Dussel's account of exteriority is not merely a negative gesture. It also offers a positive vision of philosophical method. Analectics requires not only critique, but a reorientation of philosophical attention. Rather than seeking to master the totality, the philosopher must attune themselves to what interrupts it. This involves a shift in the topology of thought: from systems that enclose to structures that remain perforated, incomplete, open to the voice of the Other. This openness is not a void but a condition of renewal. It allows philosophy to be addressed by what it cannot fully comprehend.

We might think of this as a practice of structural listening. In contrast to hermeneutic models that presume interpretive access, structural listening accepts that some discourses do not yield themselves to comprehension on familiar terms. Their opacity is not failure but fidelity—to their own historicity, to the silences imposed on them, to the asymmetries that sustain systems of thought. For Dussel, this listening is not passive; it is the beginning of critique. Vizcaíno (2021) describes a similar orientation in terms of anti-fetishist method, where fidelity to the Other's inassimilable presence becomes the ground for decolonial critique. It is how philosophy becomes accountable to what it has excluded.

Thus, opacity returns—now not as moral condition but as ontological and epistemological orientation. It marks the boundary where systems fail to account for their own exclusions, and where method must yield to encounter. This does not render philosophy impossible. On the contrary, it opens philosophy to what lies beyond its habitual form: to what is opaque, exterior, and yet still pressing.

In the next section, I turn to the recursive structures of identity and meaning, where opacity does not arise from exteriority alone but from internal self-reference—the failure of systems to coincide with themselves. There, we will see how opacity can also emerge from within: not as exclusion, but as structural recursion. Together, these views converge toward a general model of non-totalizing thought, in which opacity functions not as limit, but as method.

Recursive Identity and Structural Non-Coincidence

Opacity is not only what lies beyond the system. It also emerges from within. If Bernard Williams shows that the ethical self is partially opaque to itself, and Enrique Dussel shows that systems occlude what lies beyond them, then a third dimension of opacity comes into view when we examine recursion: the structure by which a system, subject, or process refers to itself in order to constitute or maintain identity. In recursive systems, opacity arises not because something is outside, hidden, or excluded—but because the system cannot fully coincide with its own operations. This is a structural opacity: an internal limit that arises from the dynamics of self-reference itself.

To say that a system is recursive is to say that it loops—its outputs become its inputs, its later states feed back into its earlier conditions. This is most obvious in computational systems, where recursive functions call themselves as subroutines. But recursion is not only a technical feature. It is a broader structural logic that appears in biological systems (Maturana & Varela, 1980), linguistic interpretation (Chomsky, 1965), symbolic cognition (Dennett, 1991), and indeed, in selfhood. The subject, as both a temporal and cognitive structure, is recursive: it recognizes itself, projects its future onto itself, interprets its past as part of itself. And yet in doing so, it never fully arrives at itself. Every act of self-reference produces a residue—an ungrasped difference—that cannot be resolved within the loop.

This residue is not noise or error. It is a necessary feature of recursion itself. To constitute identity through self-reference is to construct an internal delay: a point at which the system folds back but cannot complete the circuit. This is what I call structural non-coincidence. The system reflects itself but never fully aligns with the reflection. What appears as identity is always displaced—what I am refers to what I was, anticipates what I will be, recognizes what I project—but never settles into self-sameness. The recursive self is inherently skewed, not through accident but through structure.

This skew is a form of opacity. The recursive subject cannot fully bring itself into view because it is constituted by the movement of self-differentiation. There is no “center” of the system from which full transparency could be achieved. Every center would itself be a product of recursion. The desire for coincidence—between self and self, or system and ground—is structurally undermined by the very operations that make selfhood possible. As a result, opacity arises not as external resistance but as internal reflexivity.

This account deepens and radicalizes the insights of Williams and Dussel. For Williams, opacity is a condition of ethical integrity; for Dussel, it marks the exteriority that totalized systems cannot incorporate. But in both cases, opacity appears as something *outside* the dominant mode of understanding—as an ethical remainder or a political interruption. What recursion shows is that opacity does not merely interrupt systems from without; it constitutes them from within. There is no totality, not only because systems exclude, but because they cannot close over themselves. Identity, meaning, and thought are inherently incomplete—not by omission, but by structural recursion.

This has critical consequences for how we understand philosophical method. If philosophy seeks coherence, transparency, and justification, then recursion reveals the cost of such ideals. It is not that coherence is impossible in principle; it is that coherence is always bought at the price of internal simplification. To render the self-transparent is to abstract from its recursive complexity. To make a system intelligible in its entirety is to suppress its non-coincidence. Opacity, then, is not a flaw in philosophical method but a trace of what such method must suppress to sustain itself.

Recursive systems model this suppression. In computation, recursion is only tractable when it terminates or stabilizes. In language, recursive structures require interpretive heuristics to resolve ambiguity. In cognition, recursive self-models function heuristically—they do not disclose a "true self" but enable navigation through an unstable one (Metzinger, 2003). In each case, what allows the system to operate also renders it partially opaque: the system does not "know itself" completely but iterates a provisional self-understanding through feedback. The loop is not a mirror but a structuring delay—a deferral of identity.

This delay structures philosophical reflection as well. When philosophy reflects on its own foundations—when it engages in meta-philosophy—it enters a recursive loop. It attempts to account for its own operations, justify its own norms, or ground its own authority. But this gesture never lands. There is always a gap between the system that reflects, and the system reflected. Every claim to grounding becomes another element within the loop, and thus subject to the same displacement. This is not relativism. It is a recognition that foundational closure is structurally unachievable—not because we have not found the right system, but because the act of self-grounding is itself recursive.

Opacity is what emerges at the limit of this reflexivity. It is not the unknown, nor the mysterious. It is the structured non-closure of systems that cannot grasp themselves entirely. Recursion models this precisely: the point at which the output of the system loops back, creating a horizon that shifts with every cycle. In this sense, opacity is a temporal artifact—a condition produced by the system's movement through time. The recursive self is never fully present to itself because it is always in delay, never at rest. It is always approaching, never arriving.

This account aligns closely with phenomenological insights into temporality. Husserl's model of internal time-consciousness, for example, presents consciousness as a flow constituted by retention and protention —the recursive relation to just-past and just-to-come (Husserl, 1991). The self is not a point but a span, structured by its recursive relation to absence. Similarly, Merleau-Ponty describes bodily subjectivity as a form of temporal folding, where habit and anticipation structure perceptual presence (Merleau-Ponty, 2012). In both accounts, the self is not a fixed substance but a recursively spaced temporality, marked by non-coincidence. Opacity, again, is not an obstruction but a structural articulation of this spacing.

This spacing also has ethical implications. If identity is recursively constituted and structurally incomplete, then no subject is ever fully legible—to others or to themselves. This undercuts ethical models that assume transparency, consistency, or unified agency. It supports instead a model of ethical life attuned to ambiguity, delay, and partiality—not as failures to overcome, but as intrinsic conditions. Williams approached this insight obliquely; Dussel framed it in structural terms. Recursion shows how these features emerge from the very constitution of the subject.

We can now see how opacity functions at three levels: (1) the moral integrity of the agent (Williams), (2) the structural exclusion of the Other (Dussel), and (3) the recursive incompletion of self-reference (this section). Each level displaces the fantasy of totalization. Each reveals a form of opacity as structure: not a limit imposed from outside, but a necessary condition of systems that seek to know, to ground, or to represent themselves.

This reframing of opacity prepares the ground for what I will call opacity as method. If recursive systems necessarily entail non-coincidence, then the philosophical method appropriate to such systems must not seek total grasp but must remain attuned to the structural folds that make understanding possible. In the next section, I articulate this stance directly—not as a retreat from rigor, but as a commitment to a form of thought that can sustain itself without the illusion of totality.

Opacity as Method: Toward a Non-Totalizing Philosophical Practice

Opacity is often treated as a failure: a failure of knowledge, a failure of communication, a failure of structure to render itself intelligible. Within dominant philosophical traditions—especially those shaped by rationalism, transcendental grounding, or systemic coherence—opacity tends to mark a problem to be overcome. But the preceding analyses suggest a different orientation. Opacity is not merely a negative limit imposed from the outside; it is a constitutive condition that arises at every level of philosophical engagement: ethical, structural, and recursive. What emerges from this triangulation is the possibility of opacity as method—a mode of philosophical reflection that affirms incompletion, resists closure and maintains fidelity to that which thought cannot incorporate without distortion.

Opacity, as developed through Bernard Williams, protects the moral self from the demand to fully disclose its interiority. For Enrique Dussel, opacity marks the boundary between a system and what it necessarily excludes in order to function. And from the vantage of recursion, opacity emerges from within: as the structural displacement that arises when a system loops back onto itself. In each case, opacity is not incidental. It is a condition for the possibility of integrity, critique, and reflexivity. If it were eliminated, what would remain is not a purified form of thought, but a collapsed one—either morally hollow, epistemically imperial, or structurally incoherent.

To treat opacity as method is to adopt a particular stance toward philosophical systems and their limits. It is to reject the idea that philosophy must aspire to totality—whether in the form of comprehensive theories, final grounds, or universally shared criteria of justification. It is to treat non-coincidence as a condition of thinking rather than a failure to be repaired. And it is to recognize that certain forms of abstraction, systematization, or generalization will always entail a loss of remainder—a forgetting of what cannot be reduced to order.

This methodological stance has historical precedent. In continental traditions, it resonates with negative dialectics, deconstruction, and post-phenomenological critiques of presence. In Anglophone traditions, it echoes skeptical traditions and recent work in moral epistemology and feminist standpoint theory. But opacity as method differs from each of these: it does not derive from skepticism about knowledge, nor from a critique of representation, nor from a commitment to alterity as such. Rather, it emerges from a structural insight: those certain systems—ethical, political, cognitive, ontological—produce opacity through their own operations, and that to philosophize responsibly within such systems is to acknowledge that production without denying meaning.

To adopt opacity as method is not to abandon clarity. On the contrary, it demands a more disciplined clarity: one that can distinguish between precision and totalization, between conceptual articulation and metaphysical enclosure. It is to practice philosophy with an awareness of its thresholds: the point at which reflection turns back on itself and generates not insight, but delay. It is to acknowledge, as Williams does, that ethical life includes commitments that cannot be publicly justified without distortion (Williams, 1981). It is to accept, with Dussel, that some philosophical systems are blind to the conditions of their own construction (Dussel, 1988). And it is to understand, through recursion, that identity—whether personal or conceptual—is not a stable unity but a loop marked by structural misalignment (Metzinger, 2003).

Opacity, in this methodological sense, becomes a discipline of non-closure. It urges the philosopher not to evade structure, but to remain attuned to the moment when structure turns against itself. It encourages a form of conceptual patience: a willingness to think in proximity to what resists being made fully intelligible. This is not mysticism. It is a rational encounter with the fact that some concepts deform under pressure, and that such deformation is not always a flaw, but sometimes a signal.

One of the risks of opacity as method is that it may be mistaken for quietism. If totalization is impossible, why build structures at all? Why not surrender thought to the undecidable, the ineffable, the unstructured? The answer is that opacity as method does not suspend structure—it complicates its use. It affirms that structures are necessary for thought but refuses the idea that they are ever final. It commits to articulation, but not to closure. In this sense, opacity is less a retreat than an ethics of modelling: it understands that all philosophical representations are partial, recursive, and temporally delayed. It asks of the philosopher not silence, but structural humility.

This humility does not flatten argument or prevent conceptual rigor. On the contrary, it often intensifies the precision of philosophical work. To think opaquely is not to abandon clarity—it is to clarify the conditions under which clarity itself is produced. For instance, a theory of justice that pretends to cover all cases will obscure its blind spots; a theory that acknowledges its opacity—its dependence on interpretive gaps, on local judgment, on contested historical frameworks—will be more accurate precisely because it refuses universality. Methodological opacity thus enables a greater fidelity to structure, not less.

Philosophy, at its best, makes possible a sustained encounter with complexity. But that encounter requires limits. Opacity names not what we have not yet thought, but what cannot be reduced without remainder. It is the name for the residue that all systems leave behind. It is the structuring delay in recursion, the ethical silence in selfhood, the excluded voice at the boundary of reason. To incorporate opacity as method is to build systems with their thresholds intact—to leave spaces unsealed, not out of indecision, but out of recognition.

This recognition has a formal component. It shifts the criteria by which philosophical work is evaluated. Rather than asking only whether a system is complete, it also asks whether it is aware of its incompletion. Rather than seeking coherence at all costs, it allows for patterned inconsistency—for partial structures that maintain internal rigor without global closure. It allows for footnotes that signal absence, for arguments that double back, for conclusions that do not resolve but hold open. These are not marks of weakness; they are the signs of a method attuned to the non-totalizing logic of thought itself. As Vizcaíno (2021) argues, the commitment to anti-fetishist critique entails exactly this kind of methodological stance—one in which the refusal to close meaning becomes a form of decolonial rigor.

Moreover, opacity as method is not only epistemological—it is political. In resisting totalization, it also resists domination. Dussel’s critique of Eurocentric reason is not only a call for inclusion; it is a demand that philosophy account for what it structurally forgets. To apply opacity methodologically is to listen not only to what a system says, but to what it renders unsayable. It is to build frameworks that remain open to disruption—not out of fragility, but out of principled incompletion. As Sánchez-Pérez (2023) emphasizes, philosophical responsibility requires resisting the impulse to fold all difference into intelligible sameness. This is not a weakness to be overcome. It is a condition of responsibility.

Finally, opacity as method changes what it means to philosophize across traditions. It rejects the assimilation of one framework into another, or the translation of all difference into shared terms. Instead, it allows traditions to speak obliquely to one another: to resonate, to refract, to unsettle. It accepts that some concepts will not be recoverable without loss, and that this loss must be marked. In doing so, opacity becomes a practice of philosophical hospitality—not an openness that dissolves the other into the same, but one that receives the other without demand for full legibility.

As this paper has shown, opacity is not merely a topic within philosophy. It is a condition that marks philosophy's own recursive structure. To recognize this is not to despair, but to refine one's method: to philosophize from within incompletion, without seeking to erase it. Opacity, understood as a methodological stance, does not paralyze inquiry. It disciplines it—by maintaining the structural folds, recursive delays, and ethical asymmetries that make philosophy both possible and necessary.

In the concluding section, I will draw together the implications of this orientation—suggesting how opacity, as a methodological commitment, reshapes our understanding of philosophical rigor, system, and cross-traditional dialogue.

Conclusion: Philosophy Without Totalization

Opacity is not a hindrance to philosophical thought—it is its horizon. From the ethical irreducibility of the self to the structural asymmetry of system and exterior, to the recursive displacement within identity itself, opacity marks the points at which philosophy must relinquish the dream of totality. This relinquishment is not a loss but a methodological gain. It allows philosophy to think from within its own incompletion: to build without finality, to reflect without mastery, to encounter without absorption.

To adopt opacity as method is to remain attentive to the folds and failures that give systems shape. It is to construct concepts that are rigorous without being exhaustive, and to engage across traditions without the demand for full equivalence. Such a method does not unify but configures—a practice of thought shaped by delay, non-coincidence, and respect for what resists articulation. In this way, philosophy becomes not a system of knowledge but a disciplined openness to what remains outside, unassimilated, and yet structurally central.

This openness is not a passive condition but a demand. As Maldonado-Torres (2007) argues, to think from the underside of modernity is not merely to revise existing systems, but to transform the philosophical stance itself—to shift from control to receptivity, from mastery to accountability. Opacity, in this register, becomes a practice of refusal: a way of denying philosophy the authority to assimilate everything into its own terms.

Opacity, then, is not what philosophy must overcome. It is what philosophy, at its most honest, begins from—and returns to, in every act of reflection that refuses to seal the world within a closed and final form.

References

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Dennett, D. C. (1991). *Consciousness explained*. Little, Brown.

Dussel, E. (1996). *The underside of modernity: Apel, Ricoeur, Rorty, Taylor, and the philosophy of liberation* (E. Mendieta, Trans.). Humanities Press.

Dussel, E. (1988). *Ethics and Community* (R. R. Barr, Trans.). Orbis Books.

Dussel, E. (2003). *Beyond philosophy: Ethics, history, Marxism, and liberation theology*. Rowman & Littlefield.

Husserl, E. (1991). *On the phenomenology of the consciousness of internal time (1893–1917)* (J. B. Brough, Trans.). Springer.

Maldonado-Torres, Nelson. (2007). On the coloniality of being: Contributions to the development of a concept. *Cultural Studies*, 21(2–3), 240–270.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Reidel.

Merleau-Ponty, M. (2012). *Phenomenology of perception* (D. A. Landes, Trans.). Routledge.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. MIT Press.

Rodríguez Reyes, A. (2022). *Enrique Dussel y la crítica a la modernidad* [Enrique Dussel and the critique of modernity]. Heraldos Editores.

Sánchez-Pérez, J. (2023). The dangers of re-colonization: Boundaries between Latin American philosophy and Indigenous philosophy. *Comparative Philosophy, 14*(2), 119–133.

Vizcaíno, Rafael. (2021). Liberation philosophy, anti-fetishism, and decolonization. *Journal of World Philosophies*, 6(2), 61–75.

Williams, B. (1981). *Moral luck: Philosophical papers 1973–1980*. Cambridge University Press.

Williams, B. (1985). *Ethics and the limits of philosophy*. Harvard University Press.

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 249-267

Ensayo/Essay

Ser humano y conocimiento en las obras tempranas de George Siemens

Human being and knowledge in the early works of George Siemens

Freddy Varona Domínguez

Universidad de La Habana, Cuba
fvarona1960@gmail.com
https://orcid.org/0000-0002-5214-2735

DOI https://doi.org/10.48204/2805-1815.8487

INFORMACIÓN DEL ARTÍCULO

Recibido el: 4/7/2025

Aceptado el: 11/8/2025

Keywords:

Human being, knowledge, connectivism, flexibility, technology

Palabras clave:

Ser humano, conocimiento, conectivismo, flexibilidad, tecnología

ABSTRACT/RESUMEN

Abstract:

This paper reflects on the place that George Siemens gives to the human being in the conception of knowledge he expounds in his early works. The objective was to analyze the place of the human being in this conception. The idea defended is that the conception of knowledge that Siemens expounds in his early works does not suppress the human being but relegates them to lower levels of importance. One achievement of this research is having captured the overvaluation of some results of human activity above the human being; the greatest result is to note that we must be alert to the subtlety of dehumanization. Among the conclusions is that because of the fundamental importance of technology in Siemens' conception of knowledge, it can be considered a technological conception of knowledge.

Resumen:

En el presente trabajo se reflexiona en torno al lugar que George Siemens le otorga al ser humano en la concepción del conocimiento que expone en sus obras tempranas. El objetivo fue analizar el lugar del ser humano en la mencionada concepción. La idea que se defiende es que en la concepción del conocimiento que Siemens expone en las primeras obras no suprime al ser humano, pero lo relega a planos inferiores de importancia. Un logro de esta investigación es haber captado la sobrevaloración de un fruto de la actividad humana por encima del ser humano; el mayor resultado es ratificado que hay que estar alerta ante la sutileza de la deshumanización. Entre las conclusiones está que, debido a la importancia básica de la tecnología en la

	concepción de conocimiento de Siemens, esta puede calificarse como concepción tecnológica del conocimiento.

Introducción

Es común en los tiempos que corren hablar de tecnologías y se debe, en gran medida, a que se han extendido en el quehacer humano. Varios especialistas se dedican a su desarrollo con el mayor entusiasmo y apuestan por ella y los beneficios y ventajas que, de forma continua y ascendente, se obtendrán con su empleo. Otros estudiosos, en cambio, alertan respecto a las consecuencias de su utilización, cada vez más múltiples; uno de estos autores es Éric Sadin (2019), entre cuyas reflexiones está la que gira en torno a la inteligencia artificial y a impedir que el uso con el que ha sido concebida lleve a tener que redefinir al ser humano, porque ya no sea quien, con sus capacidades, ejerza su poder de acción, “sino una fuerza interpretativa y decisional que se tiene por más eficaz” (p. 146) y que lo despoje de sectores de la vida donde desde antaño ha estado posicionado. Esta relación entre el ser humano y la tecnología fue el estímulo para que se realizara la investigación filosófica, de la cual el presente trabajo constituye uno de sus resultados.

La investigación se focalizó en un autor, cuyas elaboraciones teóricas están relacionadas con las tecnologías y sus nexos con el ser humano en un universo básico de la sociedad, y no siempre justipreciado, el de la educación. George Siemens, nacido en la década de 1950, estudioso del aprendizaje y las tecnologías de información y comunicación (TIC), creó lo que algunos especialistas han dado en denominar “nueva filosofía de la educación para la era digital” (Vázquez et al., 2021, p. 55): el conectivismo, del cual, una parte considerable ha desarrollado junto a su colega y coterráneo Stephen Downes (2012; 2022), quien lo define como la tesis de que el conocimiento se distribuye a través de una red de conexiones y que el aprendizaje consiste en la habilidad de construirlas y atravesarlas. Aun cuando es notorio el trabajo conjunto de ambos, las ideas de Siemens, como fundacionales, merecen una atención especial y así se hizo, en particular con las de sus primeras obras.

Las elaboraciones teóricas de Siemens tienen en su centro de atención un asunto histórico de la filosofía: el conocimiento, en torno al cual reflexiona en el marco académico, de ahí que esta ocasión se estudió desde la Filosofía de la Educación. Los antecedentes de esta investigación están en los estudios realizados, desde dicha especialidad, a los cambios en la educación superior resultantes de la utilización de la tecnología y de la influencia de concepciones epistemológicas actuales; así, en este caso, el primer paso fue profundizar en las elaboraciones teóricas de Siemens y en los criterios acerca de ellas.

Con el paso iniciador emergió la diversidad de razonamientos en torno a las ideas del canadiense, entre ellos, los siguientes: son valiosas para la realización de las clases en línea (Nivela et al., 2022); apuntan hacia las transformaciones a realizarse en el universo educacional (Cerquera & Álvarez, 2021), a contrarrestar la rápida obsolescencia de los conocimientos (Antonelli, 2024) y a aprovechar los vínculos entre las áreas cognoscitivas y los contextos culturales (Llorente et al., 2024). Pero apareció una posición opuesta, desde donde se duda de su condición de teoría científica, porque se limita a describir el aprendizaje en ambientes tecnológicos (Islas, 2021) y porque carece de un fundamento consolidado en la práctica, de una metodología y de precisión de los objetivos, métodos, valores y aportes al aprendizaje (Suárez-Guerrero et al, 2022).

La anterior malla de criterios mostró que en los estudios sobre Siemens no sobresale la perspectiva filosófica, por ende, el presente estudio se desplegó desde la Filosofía de la Educación, guiada por la pregunta filosófica siguiente: ¿Qué lugar le da Siemens al ser humano en la concepción del conocimiento que expone en las obras tempranas? Para responderla, el objetivo propuesto fue: analizar el lugar del ser humano en la concepción del conocimiento de Siemens en las obras tempranas. La idea que se defiende es la que en la concepción del conocimiento que Siemens expone en las primeras obras no suprime al ser humano, solo lo relega a planos inferiores de importancia, que es en sí una muestra de deshumanización.

La estrategia argumentativo-metodológica seguida fue tomar como bibliografía básica los primeros textos de Siemens; se priorizaron los existentes en español, ante todo *Conociendo el conocimiento* (Siemens, 2010), por su valor teórico; también se estudiaron algunos de Downes (2012; 2022) y de otros autores, de los cuales se utilizaron los más enriquecedores. Con las fuentes bibliográficas se utilizó el método de revisión crítica y análisis del discurso. En la investigación se utilizó la metodología documental de estudio crítico de textos (Priscal, 2021). El punto de partida epistemológico fue la recomendación de Feyerabend (1975) de realizar el análisis sin la interferencia de reglas preestablecidas, que se complementa con la de Siemens (2010) de “no crear nuestros filtros por adelantado” (p. 20); así, la guía metodológica fue entender las reflexiones del canadiense desde ellas mismas. El trabajo está dividido en cuatro secciones.

La definición no deseada: punto inicial de la concepción del conocimiento

La primera acción que realiza Siemens (2010) en sus elaboraciones teóricas es atacar la visión según la cual el conocimiento es “estático, organizado y definido por expertos” (p. 3) y, a la vez, fortalecer la mirada que posibilita concebirlo de modo dinámico y multifacético, propósito tenido también por varios estudiosos, sobre todo en los últimos años, entre ellos Morin (1999), quien a finales del siglo pasado señaló que debía repensarse y “problematizarse” (p. 34). El canadiense decide ejecutar el ataque mediante la caracterización, debido a la resistencia a usar definiciones, sobre todo las que llama totalitarias.

Esas definiciones no son para él las que abarcan el objeto en su totalidad, sino las que llegan a ser consideradas únicas; las denomina precisas, “verdadera” (p. 17) porque llegan a ser vistas como eternas. Sostiene el siguiente principio: “Cuanto más precisa es una definición, menos aplicable es a múltiples situaciones” (p. 17). Sin embargo, aclara que lo específico puede funcionar en un contexto amplio, y que en tal circunstancia hay que ver el conocimiento por lo que es por sí mismo, buscar y hallar la especificidad y actuar en correspondencia.

La posición anterior se debe a que piensa que ese tipo de definiciones cierra el paso al conocimiento, porque son vistas como definitivas y eso es razón para que se desechan otras percepciones, valiosas, que, incluso, pueden ser superiores. Ese parecer del canadiense hace pensar que no justiprecia la frecuente explosión de las fronteras de las definiciones a causa del crecimiento del conocimiento, ni que la pretensión de abarcar el fenómeno en su mayor amplitud, propia de la filosofía, no por obligación tiene que ser estática, ni obstaculizar los conceptos específicos.

Siemens (2010) insta a que se tome conciencia de que “el conocimiento ha cambiado” (p. XV) y que ha transitado de categorizaciones y jerarquías, a un nivel diferente, constituido por redes y ecologías; de estas últimas se hablará más adelante.

Insiste en que la caracterización del conocimiento ha de hacerse no a partir de puntos de vista personales, sino de las características del conocimiento en sí. Este afán de objetividad se debe a que renuncia a la utilización de filtros previos, capaces de enturbiar la visión y así, al estilo de Morin (1999a), se propone evitar los errores e ilusiones impregnados en las ideas, y protegerse de ambos.

Siemens (2010) escribe sobre el conocimiento convencido de que la faena cognoscitiva se puede realizar de muchas formas, ya sea “como una entidad y como un proceso, como una secuencia de continuos: tipo, nivel, y aplicación, implícito, explícito, tácito, procedimental, declarativo, inductivo, deductivo, cualitativo y cuantitativo” (p. 14), que en cada caso es diferente y que es un error mirarlo a través de lentes monocromáticos.

La visión policromática del canadiense se corresponde con el modo de concebir el conocimiento como una organización. Es aquí donde destaca al ser humano, porque sostiene que la finalidad a tenerse con el conocimiento no es ocupar la mente, sino abrirla, y si tradicionalmente se había perseguido que fuera una organización que se caracterizara por la claridad, la estabilidad y la distribución en jerarquías y una especie de contenedores o reservorios, ahora lo que ve a propósito del conocimiento son redes dinámicas tecnológicas. Las redes y las conexiones que se logran con ellas son el centro de atención del canadiense en cuanto al conocimiento; el ser humano queda en segundo plano.

Aunque muy temprano en las elaboraciones teóricas se declara no simpatizante de las definiciones y reúsa elaborar una acerca del conocimiento, Siemens (2010) lo define; no solo una vez, ni de un modo único; así demuestra que no pudo prescindir de esa herramienta tan valiosa en el quehacer intelectual. Para definirlo, se apoya en la relación entre la información y los datos, porque considera que el conocimiento es “la codificación de información o datos de una determinada forma” (p. 21); o sea, lo ve como diversos tonos de información. Pero esta función la puede realizar algo que no es el ser humano, con más rapidez y mayor precisión; por esa razón, relega al ser humano a un segundo nivel. No obstante, mediante la información, es que el ser humano retorna a las reflexiones del canadiense, porque ella se transforma en conocimiento cuando se interioriza y este “dato” el canadiense no puede ignorarlo.

Pero también define el conocimiento de otro modo, y lo hace con una considerable amplitud, en tanto afirma que “en realidad es más una corriente” (p. 52), ya que es algo que corre; es un fluido indetenible y con conexiones. Esta idea la completa con una puntualización: Conocer es “estar conectado” (p. 51). Pero, ¿quién se conecta? La primera respuesta es que es el ser humano, y, por tanto, es él quien realiza el acto de conocer. Pero no es así. El canadiense destaca que estar conectado no es propiedad exclusiva de los humanos.

Enfatiza que el conocimiento existe “en dispositivos no-humanos” (Siemens, 2010, p. 149) y que está ligado, básica y esencialmente a la conexión, que es donde ve su existencia, la cual está en continua transformación. No obstante, no niega que existe en los seres humanos y sus relaciones; e insiste que el flujo de información es en múltiples sentidos. Ahora bien, asegura que mediante las redes cada persona descubre el valor del conocimiento y de su fluidez, y que esta experiencia depende de la “«madurez epistémica» del individuo” (Siemens & Weller, 2011, p. 158); es decir, que el conocimiento es una actividad no humana y humana, con gran carga objetiva y subjetiva, pues no pierde oportunidad para insistir que vale por sí mismo.

La visión del conocimiento que expresa Siemens (2010) es dinámica, como las conexiones; y como ellas, es, a su vez, multidireccional, aunque se mantiene ligada a algo poseedor de mucha importancia en su concepción: el contexto, que es lo que determina que “un punto de vista puede ser el más adecuado” (p. 13) y que la definición no debe ser rígida, antes bien, flexible. He aquí dos aspectos básicos de su concepción del conocimiento. Puede compartirse el criterio de Villalba (2024) en cuanto a que el canadiense concibe “como inestable, volátil, incontrolable y en constante expansión, lo cual significaría fuera de control de una persona y que puede permanecer en redes externas, sean comunidades, dispositivos digitales, entre otros, que además también están en constante cambio” (p. 5941).

Contexto y flexibilidad: dos categorías básicas para el conocimiento

Estas dos categorías tienen gran valor para Siemens (2010). Le otorga importancia básica al marco contextual y a la relación que se teje en él y a propósito de él. Esta posición de principio recuerda a Morin (1999a) quien asevera: “Hay que ubicar las informaciones y los elementos en su contexto para que adquieran sentido” (p. 14), para que tengan concreción y con ella, validez. La valía que Siemens (2010) le concede al contexto es tal, que incluso sostiene que es donde se evidencia lo que es el conocimiento, su naturaleza y la definición que se necesita; incluso, parafraseando a Wittgenstein, asegura que “la mayor parte del proceso del conocimiento consiste en un juego de contextos” (p. 16). Sobre esta base, rechaza las definiciones apriorísticas, característica que puede entenderse, al estilo de Antolínez (2023), como protección ante la extrapolación mecánica de ideas.

La significación que el contexto tiene para Siemens (2007) conduce a que tome consistencia otra idea significativa: la importancia básica del *saber dónde,* como complemento del *saber cómo* y *el saber qué*. No basta determinar qué es lo que se quiere, ni cómo hallarlo; el propósito cognoscitivo precisa tener diafanidad respecto a “dónde encontrar el conocimiento requerido” (p. 2). Esta labor incluye, como apuntan Sobarzo-Ruiz et al. (2023), la posesión del conocimiento que permita ubicarse en el lugar adecuado para buscar la información necesaria y hallarla; esto es, identificar las conexiones capaces de propiciar el conocimiento deseado (Gómez et al, 2021). Saber ubicarse para encontrar el conocimiento es lo más valioso y decisivo para el canadiense, más que cómo asimilarlo y más, incluso, que el propio conocimiento; pero no puntualiza quién debe saber ubicarse. Puede pensarse que es el ser humano en general; pero este principio tan común cae en dudas ante el reconocimiento de que la capacidad cognoscitiva no es exclusiva del ser humano.

Siemens (2010), a partir de que solo se propone caracterizar, expresa que un contexto incluye elementos: el territorio donde tiene lugar algo y el ambiente que lo rodea, dado por las emociones, experiencias y creencias, entendidas como algo objetivo; así como otros aspectos: política, ideología, historia, cultura, enfoques, perspectivas, todo lo cual forma una malla heterogénea que influye sobre la cognición, similar a como la concibe Morin (1999; 2004; 2005). Así, atiende el debate como un asunto epistemológico, tanto el modo de realización, como el ambiente donde se desarrolla, en estrecha relación con la cultura y las ideologías, aunque a estas últimas las evita, porque quiere impedir que el debate esté de antemano “esencialmente fijado” (Siemens, 2010, p. 62) y que se encamine a la proyección de visiones acerca del mundo.

Siemens (2010) le otorga valor al espacio, porque en él tienen lugar las conversaciones y los encuentros cognoscitivos y es donde están las instituciones vinculadas al conocimiento, entre ellas destaca Internet, pero el espacio de verdadera importancia para él es el de las redes y las ecologías, que son las que posibilitan el flujo de conocimiento.

Esas últimas, las ecologías, son para él “un entorno en el que se comparte conocimiento” (Siemens, 2010, p. 86), son, a la vez, “modelos sensibles a la adaptación, que se ajustan y reaccionan a los cambios” (Siemens, 2011, p. 21), por ello, se destacan, según su comprensión, por la diversidad, la pluralidad y la flexibilidad, cualidades que posibilitan la gestión del conocimiento desde diferentes perspectivas, pero, sobre todo, con apertura, y, que, a la vez, sean un impulso hacia la innovación. Toda ecología se caracteriza por ser informal, no estructurada, con riqueza de herramientas, pero también por la descentralización, la simplicidad de ideas y enfoques en su utilización y por la “alta tolerancia a la experimentación y al error” (Siemens, 2010, p. 88), pues así se estimula el crecimiento del conocimiento y la innovación: para todo esta es importante el espíritu inquisitivo y son imprescindibles la apertura mental y la tolerancia, que son cualidades humanas; por tanto, el ser humano ha de estar presente. Pero de sus reflexiones se infiere que tales cualidades están más allá del ser humano pues también se les puede hallar en las redes, es decir, en el entramado que forman las redes y en las conexiones que con ellas llegan a crearse.

En ese entramado es donde germina el espíritu de la flexibilidad, una característica de las elaboraciones teóricas de Siemens que tiene gran implicación en el ámbito académico (Haris et al., 2023), ante todo por su significación epistemológica. En la lengua española la palabra *flexibilidad* tiene entre sus significados el de fácil adaptabilidad a las características del otro y que puede variar en dependencia de las circunstancias. A partir de estas acepciones se usa para referir el pensamiento abierto y dinámico, apto para aceptar las nuevas posibilidades (Hayes, 2020, p. 38). De modo similar sucede cuando la alusión es a la flexibilidad mental, que se trata de la manera de pensar poseedora de la capacidad de no frenarse ni cerrarse a las novedades, probabilidades, contingencias, alternativas, ni a los disímiles puntos de vista; tampoco a la diversidad.

Más cercana a Siemens es otra categoría con significado próximo a los anteriores, pero más propia del universo epistemológico: flexibilidad cognitiva; con ella se hace referencia al dinamismo en el proceso del conocimiento, a la apertura al cambio y a la adaptabilidad en el mismo. La han desarrollado teóricamente varios autores, entre ellos Paba-Barbosa et al (2019), quienes la asocian más a la movilidad del comportamiento que a la apertura, y apuntan que en la bibliografía revisada por ellos se concibe como un proceso constituido por componentes: diversidad de ideas, respuestas alternativas, modificación de planes y conductas, así como la capacidad para alternar diferentes respuestas, identificar errores y modificar conductas mediante estrategias alternativas y distintos focos. Criterios similares tienen Bernal-Ruiz et al (2023) porque recalcan la variación de las circunstancias y la adaptación a las mismas, así como la disposición de afrontar con coherencia las novedades, crear alternativas y aprender de los errores. Estos juicios están relacionados con Spiro & Jehng (1990) quienes a finales del siglo pasado la habían definido como la capacidad de reestructurar el conocimiento en respuesta adaptativa a demandas situacionales nuevas.

En toda la obra Siemens pide flexibilidad, aunque casi de modo silencioso, pero con pretensión diáfana: satisfacer las exigencias cognoscitivas actuales, ligadas a las redes, la conectividad y la velocidad que se ha logrado y aumentará, para lo cual hay que vencer las costumbres epistemológicas que se han convertido en trabas y en fuerzas capaces de lentificar. El derrotero de sus reflexiones es el futuro, donde será más importante el conocimiento.

Tipos de conocimientos a partir de las conexiones

Saber sobre, saber ser, saber hacer, saber dónde, saber transformar, son los tipos de conocimiento que Siemens (2010) menciona. Los dos primeros los asocia a las estructuras de almacenamiento de conocimiento existentes: revistas, libros, bibliotecas, museos, pero de los otros tres puntualiza que "están más allá de estas perspectivas-contenedor" (p. 10) y pide brindarles mayor atención. Sostiene que esta se puede viabilizar mediante dos categorías: conocimiento conectivo y conocimiento distribuido, que bien vistas, también son tipos.

Para tratar este tema es imprescindible destacar una consideración de Siemens (2010) respecto al conocimiento: "Lo que antes era el medio, ahora es el fin" (p. 3), o sea, que, según su visión, en la actualidad es más importante aquello que era visto como instrumento, vía, contexto o entorno donde existía el conocimiento o mediante el cual circulaba y que, por tal razón, era menos atendido que el contenido, es decir, el conocimiento en sí. El medio que él refiere son las redes y las conexiones posibles por ellas. De aquí su aseveración de que en la actualidad el conocimiento se define "por las conexiones" (p. 15).

En las elaboraciones teóricas de Siemens (2010), la conexión tiene lugar entre entidades; sucede cuando, por lo menos, una propiedad de una de las entidades conduce a alguna propiedad de la otra o se convierte en propiedad de ella. La esencia de la relación está en las propiedades, ya sea porque se crean nexos entre ellas o porque una de las cualidades de una propiedad pasa a ser cualidad de la otra entidad. En los nexos se crea un tipo de conocimiento que denomina conectivo, porque es resultado de las conexiones, las cuales, a la vez, son frutos de las redes y de los nodos que las forman, los cuales propician la conexión con nuevas fuentes de información, que será procesada por la red cognitiva humana interna, mediante la cual cada ser humano se conecta con el mundo externo para conocerlo.

De lo anterior se infiere que reconoce la presencia humana y su importancia, en este quehacer, y deja entrever que, para él, el ser humano es también una red, con conexiones. Tal faena lleva implícita otras tareas en diferentes niveles cognoscitivos: observación, conceptualización, descripción, explicación, valoración, transformación, que ha de realizar, ante todo, pero no únicamente, el ser humano. Desde esta perspectiva, la clave para el conocimiento no es la red interna cognitiva humana, sino la externa, o sea, la tecnológica. Por tanto, el conocimiento conectivo es "el producto de una forma particular de codificar información o de procesar datos en una conexión" (Peña, 2019, p. 3); es el resultado de la conexión y la esencia de la propia conexión, cuya razón de ser es la creación de este tipo de conocimiento, cuya esencia es la conexión que se logra mediante las redes, no ante todo las humanas, sino las tecnológicas.

Con la categoría conocimiento conectivo, Siemens da denominación al conocimiento que no se encuentra en una sola fuente o lugar, porque existe en las conexiones, diseminado en ellas y en las redes, que es donde se difunde y es posible adquirirlo. Esta reflexión tiene una valía especial en la actualidad, cuando se extiende el contexto digital, donde, como expresa Gallego (2020) se observa una descentralización cognoscitiva en tanto se han alterado los roles de los actores del proceso del conocimiento, pues la certeza y legitimación ya no están de manera exclusiva en el universo de los científicos, ni en los lugares donde históricamente se concentraban: universidades y centros de investigación; además, los conocimientos circulan por diversos canales, “disponibles a través de una amplia variedad de fuentes y medios” (Gallego, 2023, p. 535) y al alcance de personas que tienen posibilidades para aprender y opinar. Solo si se entiende esta situación, se puede comprender que el fin de Siemens con la categoría conocimiento conectivo, más que referir la asimilación del conocimiento, lo que nombra es la existencia, elaboración y difusión del conocimiento mediante las conexiones que posibilitan las redes.

De las redes de conocimiento Siemens (2010) menciona cuatro características: diversidad, autonomía, interactividad y apertura. No las explica; de ellas solo emite preguntas, quizás siguiendo la sugerencia que Morin expresa más de una vez, de darle mayor importancia a las preguntas que a las respuestas. Con las interrogaciones se puede suponer que alude lo esencial de cada una. La primera y la última de esas características, es decir, diversidad y apertura, sugieren flexibilidad, porque esta es una condición vital para la diversidad y la apertura; la rigidez y el esquematismo no las permiten.

La diversidad, la apertura y con ellas, la flexibilidad, son el modo de existencia del conocimiento entendido como constelación de conexiones mediante las redes tecnológicas. De tal suerte, se fundamenta en la diversidad de opiniones provenientes de los diferentes nodos conectados a la red, que debe ser moldeable para permitir la creación de más nodos y más interconexiones a fin de que el conocimiento se incremente; el hecho de percibir las “conexiones entre campos, ideas y conceptos es una habilidad central” (Peña, 2022, p. 38), pero Siemens no llama la atención a todo tipo de conexiones, sino a las tecnológicas.

Reacio a las definiciones y propenso a la caracterización, Siemens (2010) usa otra categoría próxima a la anterior: conocimiento distribuido. No refiere el conocimiento que está dentro de una persona, ni vive solo en las personas, sino “como una función de elementos distribuidos a través de un sistema” (p. 44), que circula a creciente velocidad y se multiplica muy rápido. Así hace pensar que existe en muchos lugares, que es ubicuo, como Dios, porque está en todos los sitios al mismo tiempo. Con esta categoría queda en evidencia que el conocimiento tiene un segundo plano respecto a las conexiones. Enfatiza que esta situación se observa con mayor claridad en las escuelas, donde los estudiantes se sienten estimulados a encontrar más valor en las perspectivas agregadas o combinadas, se convierten en creadores de contenidos y el aprendizaje se transforma en continuo, exploratorio y carente del control de alguien específico.

La categoría conocimiento distribuido se relaciona con la magnitud y diversidad del conocimiento en la actualidad, así como con los problemas y soluciones asociados a él, que no cabe en la mente de un individuo y está diseminado en las redes. Puede pensarse que con la categoría el canadiense no pretende significar que el conocimiento existe de modo extendido por doquier, sino que "está en las relaciones entre las personas que participan en una actividad, las herramientas que utilizan y las condiciones materiales del entorno en el que la acción tiene lugar" (Santamaría, 2010, p. viii), pero de los textos puede entenderse que el conocimiento distribuido, que es en sí una metáfora, alude no al modo tradicional de entender la concentración de conocimiento, sino a la que tiene lugar en las redes tecnológicas y las conexiones existentes gracias a ellas, cuya particularidad es que en ellas el conocimiento se concentra de manera explayada, gracias a la capacidad de multiplicarse y trasladarse a una velocidad creciente, con la cual se forma la idea de ubicuidad.

En los textos del canadiense hay una invitación a adaptarse a la velocidad que se alcanza en estos tiempos y considera meritoria la capacidad de adaptabilidad, tanto, que lo más importante no es el conocimiento, ni la precisión que posea, sino la capacidad de adaptarse a las nuevas circunstancias marcadas por la celeridad y la actualización; valora mucho a quien está actualizado, porque pudo adaptarse a la época actual, signada por la rapidez y porque el conocimiento no solo lo genera y almacena el ser humano, sino también las redes, que es donde se cambia y actualiza. Es evidente el desplazamiento del ser humano a la tecnología: derrotero de las ideas de Siemens.

Esa posición la expresa con las categorías conocimiento conectivo y distribuido, que se conforman y consolidan en las conexiones y las redes tecnológicas, las cuales, valga la reiteración, son las protagonistas verdaderas de sus reflexiones; aunque considera que el ser humano es quien debe decidir cuál conocimiento está actualizado u obsoleto y cuál tomar o desechar, así como distinguir entre el conocimiento personal y el que está dentro de un ámbito particular, todo lo cual promete mucho debate en el futuro (Siemens & Weller, 2011). No han de confundirse el conocimiento que posee una persona y el saber, entendido como área de acumulación, procesamiento y utilización de determinada información; ni olvidarse que existen personas que no poseen riquezas cognoscitivas y son enormes consumidoras de información.

El conocimiento hoy está relacionado más que nunca antes con la celeridad de la vida actual, característica que provoca varias consideraciones, como la de Campillo (2023), quien se muestra preocupado ante el choque entre dicha aceleración y "los límites biofísicos de nuestro planeta" (p. 4). Tal velocidad estimula a Siemens (2010) de otra manera y por ella distingue dos tipos de conocimiento: duro y blando.

Del primero lo más significativo que señala Siemens (2010) es que es considerado un producto, es decir, algo concluido, listo para consumir, avalado por un grupo de expertos, quienes tenían el poder de dictaminar y legitimar. Este modo de concebir el conocimiento lo ve como propio de las épocas cuando la lentitud de los cambios era una característica consustancial. Significativo es el adjetivo que emplea: duro, que puede conducir a que se interprete equivalente de profundidad y madurez, pero también de eternidad, rigidez, cierre. Ahora bien, la generalización que realizó el canadiense, evidencia que su reflexión fue esquemática y no dialéctica.

La categoría conocimiento blando le sirve para subrayar los tiempos que corren, con la velocidad que le es propia y con el continuo y acelerado cambio, cuya tendencia es no solo a mantener la transformación, sino a que sea mayor y más rápida, por lo cual aumenta la posibilidad de que el conocimiento sea corregido o reemplazado por otro nuevo, con una creciente rapidez. Por ello, se incrementa la cantidad de conocimiento que se pierde o desestima, sin que se haya podido procesar o que se haya procesado a medias, sin haber extraído todo su valor.

No niega el conocimiento duro, tampoco el blando, ni toma partido por uno u otro; la pretensión de Siemens es llegar a tener procesos capaces de gestionarlos como un continuo, no como opuestos, ni siquiera como distintos. Con esta posición refuerza su insistencia por la flexibilidad. Sobre esta base, Siemens (2010) llama la atención hacia preguntas epistemológicas que han existido a través de varios siglos, como quién crea el conocimiento, cómo llega a estructurarse, de qué modo se distribuye, cómo y quién lo legitima y de cuál manera se adquiere e implementa, y destaca que están cambiando los proceso para responder estas interrogantes, ante todo porque las tecnologías, en particular las redes, abren nuevas posibilidades cognoscitivas, como el acceso directo a los especialistas, la publicación de criterios y el despliegue de debates. Estas reflexiones son muestras de que reconoce el lugar y papel singulares del ser humano en el quehacer cognoscitivo, aunque para él las tecnologías son las protagonistas.

Individuos y grupos de personas en la concepción de conocimiento

En los razonamientos del canadiense hay dos ideas trascendentes: el conocimiento no es una posesión individual que radica en la cabeza de una persona, y, es imposible que alguien posea todo el acervo cognoscitivo alcanzado. De aquí asevera que los diversos conocimientos están almacenados más allá del individuo, en los grupos de personas, pero, sobre todo, en la tecnología.

Sobre esa base sostiene la relación epistemológica entre el individuo y los grupos de personas, que constituye para él una posición de principio. En un momento de su quehacer intelectual expresa que el punto de partida de sus elaboraciones teóricas es el individuo, pero puntualiza que “el conocimiento personal se compone de una red, la cual alimenta a organizaciones e instituciones, las que a su vez retroalimentan a la red, proveyendo nuevo aprendizaje para los individuos” (Siemens, 2007, p. 7); esta idea la fortalece años después, y asegura que “el conocimiento depende de los individuos, pero reside en el colectivo” (Siemens, 2010, p. 14). Con esta aseveración intenta delimitar la actividad cognoscitiva de un individuo y el conocimiento como acervo cultural.

Es evidente el rol significativo que tienen para el canadiense los vínculos entre el individuo y los grupos de personas, ya sean pequeños grupos específicos o la totalidad que forma la sociedad, pues cada uno de ellos enriquece al conocimiento y propicia su actualización. Pero esto no impide que enfatice la tecnología y, con ella, el intercambio de conocimiento rápido, el diálogo en niveles superiores que se logra, la comunicación que se alcanza y la colaboración que puede establecerse. Esta última es consecuente con la idea del conocimiento blando: abierto a muchas personas mediante las redes, donde se es posible opinar sobre él y hacerle cambios.

A partir de la posición anterior, el canadiense pone en el tapete teórico el asunto de las relaciones entre dos grupos de personas: por un lado, los expertos, y, por el otro, los interesados no especialistas y que conocen poco el tema, con lo cual, al mismo tiempo, saca a relucir el polémico tema de la legitimación del conocimiento, que va dejando de ser privativo de determinadas personas, gracias a las redes y las posibilidades cognoscitivas que crea. Pero aquí hay un detalle que no ha de menospreciarse y es que el conocimiento, sobre todo el científico, no es asunto de democracia; a la verdad no se llega mediante votación y esta vía, a todas luces, no es la adecuada para llevar el conocimiento a alturas superiores; además, no es conveniente que se diluya el papel de los especialistas; no ha sido fácil llegar al nivel que han llegado, sobre todo quienes han mostrado mayor cantidad de cosechas.

La posición anterior no entra en contradicción con las posibilidades que ha de tener cada ser humano de desarrollar sus capacidades y llevarlas a los grados superiores que las mismas permitan; como tampoco contradice que "la «sabiduría de las multitudes» solo funciona cuando cada uno de los miembros del colectivo aporta una perspectiva propia al espacio" (Siemens, 2010, p. 56). Hacia el conocimiento no debe haber puertas que cierren el acceso y, de hecho, los caminos han de estar abiertos a cada ser humano, con su individualidad, lo cual es una condición para la manifestación plena de la creatividad y con ella, para el bienestar de la sociedad.

Siemens (2010) muestra ser consciente de ello cuando afirma lo siguiente: "La colectividad requiere de voces individuales combinadas, no solapadas" (p. 56). En el intercambio social está el terreno fértil del crecimiento humano, además, estos tiempos no son los del pensamiento único; hoy cada individuo debe poder expresarse en la más completa libertad y que se le respete el derecho a pensar y decir lo que piensa, aunque esté equivocado, así como a ser quien desea ser, con sus especificidades.

Conclusiones

Siemens muestra una concepción tecnológica del conocimiento. La esencia de la base epistemológica de las elaboraciones teóricas de Siemens consiste en el protagonismo de las tecnologías y la relegación del ser humano a un nivel inferior al que le otorga a uno de los frutos del quehacer humano: la tecnología. Esta es la base de sus elaboraciones teóricas.

El papel protagónico que le otorga a las redes tecnológicas hace pensar en la necesidad de velar por que el ser humano no sea relegado a planos inferiores. El entusiasmo por la tecnología no puede hacer perder el derrotero de beneficiar al ser humano y contribuir a su desarrollo multilateral, que es incompatible con cuanto signifique suplantarlo o minimizarlo.

En el contexto Siemens tiene en cuenta la creatividad humana, y al hablar de la flexibilidad de las redes y las conexiones su intención es beneficiar a quienes las usen; por tanto, la importancia de ambos aspectos posibilita la emergencia del ser humano en la concepción de referencia.

En las obras consultadas, Siemens no hace una delimitación estricta entre el conocimiento como acervo cultural de la humanidad y como adquisición de saber por los humanos. Con las categorías conocimiento conectivo y conocimiento distribuido refiere la existencia de la primera variante, es decir, acervo cultural, de carácter objetivo, donde les otorga primacía a las tecnologías. Esta carencia se hace notar en la relación que establece entre el individuo y los grupos de personas.

Un logro de esta investigación es haber captado una sutileza de la deshumanización, la que tiene lugar en la sobrevaloración de algún resultado de la actividad humana por encima del propio ser humano.

El mayor resultado de este estudio es notar, una vez más, que hay que estar alerta ante la sutileza de la deshumanización.

Referencias bibliográficas

Antolínez, D. (2023) Releyendo a Latour: ontología, hibridación y diferencias. *Isegoría*, (69), pp. 1-13, e14. https://doi.org/10.3989/isegoria.2023.69.14

Antonelli, A. (2024). Foucault y Deleuze: convergencias epistemológicas y políticas. *Tópicos, Revista de Filosofía*, (68), pp. 123-151 http://doi.org/10.21555/top.v680.2459

Bernal-Ruiz, F., Ahumada, Y., Castillo, M., Castillo, C., López, Y. & Rojas, A. (2023). Flexibilidad cognitiva e inhibición como predictores de las competencias matemáticas tempranas en preescolares. *Perspectiva Educacional. Formación de Profesores*, 62(3), pp. 75-98 DOI: https://doi.org/10.4151/07189729-Vol.62-Iss.3-Art.1369

Campillo, A. (2023). La humanidad terrestre. Una filosofía del Antropoceno. *Isegoría*, (69), pp. 1-16, e25. https://doi.org/10.3989/isegoria.2023.69.25

Cerquera, C. & Álvarez, J. H. (2021). El enfoque conectivo de la internacionalización del currículo en la práctica pedagógica docente. En R. Quiroz y J. Echeverri (Coor.). *Sentidos, enfoques y perspectivas de la investigación en educación en tiempos de incertidumbre* (pp. 99-108). Editorial Universidad Pontificia Bolivariana-Universidad de Antioquia.

Downes, S. (2012). *Una introducción al conocimiento conectivo*. https://docs.google.com/document/d/17Id09r4pem_XIG5zz8uptveECkWqw0R7LjUjL0wGFh0/edit?pli=1

Downes, S. (2022). Connectivism [Conectivismo]. *Asian Journal of Distance Education*, 17(1), 58-87. https://doi.org/10.5281/zenodo.6173510

Feyerabend, P. (1975). *Tratado Contra el Método*. Madrid: Editorial Tecnos.

Gallego, R., A. (2020). Reflexiones epistemológicas para una nueva era. *Revista Luciérnaga Comunicación*, 12(23), 1-9. https://doi.org/10.33571/revistaluciernaga.v12n23a

Gallego, R. A., (2023). Generación del conocimiento en un ecosistema digital de aprendizaje. En D. Moya (coord.). *Convergencia mediática: nuevos escenarios, nuevas perspectivas* (pp. 530-540). Dykinson.

Gómez, A., Restrepo, E., & Becerra, R. A. (2021). Educational Foundations for the Creation and Production of Open Access Educational Resources (REA) [Fundamentos educativos para la creación y producción de recursos educativos de acceso abierto]. *Anagramas Rumbos y Sentidos de la Comunicación*, 19(38), 35-68. https://doi.org/10.22395/angr.v19n38a3

Haris, M., Husin, S. F., Rosli, R., & Rahmat, N. H. (2023). Is there Connectivism in Online Engagement? [¿Existe conectivismo en la interacción en línea?]. *International Journal of Academic Research in Business and Social Sciences*, 13(8), 1412-1427. http://dx.doi.org/10.6007/IJARBSS/v13-i8/17736

Hayes, S. C. (2020). *Una mente liberada*. Planeta, S. A.

Islas, C. (2021). Conectivismo y neuroeducación: transdisciplinas para la formación en la era digital. *Ciencia Ergo-sum*, 28(1), 1-14. http://doi.org/10.30878/ces.v28n1a11

Llorente, Á, Orozco, M. L. Sanz, M. (2024). La educación culturalmente receptiva: análisis educativo del tercer espacio. *Teoría de la Educación. Revista Interuniversitaria*, 36(1), 73-93. https://doi.org/10.14201/teri.31408

Morin, E. (1999). *El Método III. El conocimiento del conocimiento*. Ediciones Cátedra.

Morin, E. (1999a). *Los siete saberes necesarios para la educación del futuro*. UNESCO.

Morin, E. (2004). La epistemología de la complejidad. *Gazeta de Antropología*. http://www.ugr.es/~pwlac/G20_02Edgar_Morin.html

Morin, E. (2005). *Introducción al pensamiento complejo*. Gedisa

Nivela, M. A., Otero, O. F., Tenesaca, C. A., & Morales, E. F. (2022). Plataformas virtuales en la Educación Superior. Una visión conectivista. *Revista de Ciencias Humanas,* *Teoría Social y Pensamiento Crítico EXTRA*, (1), 155-175 http://doi.org/10.5281/zenodo.6551077

Paba-Barbosa, C., Paba-Argote Z., Barrero-Toncel, V. (2019). Relación entre comprensión lectora y flexibilidad cognitiva en estudiantes de una universidad pública. *Duazary*, 16(2), 87-104. Doi: https://doi.org/10.21676/2389783X.2944

Peña, D. (2022). Pro y limitaciones del conectivismo de acuerdo con el sistema de gestión educativa en el Ecuador. *Ciencia y Educación*, 3(11), 33-45 https://www.cienciayeducacion.com/index.php/journal/article/view/166/294

Peña, M. (2019). Internacionalización conectiva del currículo. Fundamentos epistemológicos y metodológicos. *Espacio*, 40(26), 2-9.

Priscal, R. (2021). La subversión tecnológica de la vida cotidiana. Un análisis desde el pensamiento complejo de Morin. *Ciencia Latina. Revista Científica Multidisciplinar*, 5 (1), 436-458 https://doi.org/10.37811/cl_rcm.v5i1.242

Sadin, É. (2019). La inteligencia artificial: el superyó del siglo XXI. *Nueva Sociedad*, (279), 141-148 www.nuso.org

Santamaría, F. (2010). Introducción. La era conectiva: por el desorden natural de los artefactos y nodos, en *Conociendo el conocimiento*. Traducción de Emilio Quintana, David Vidal, Lola Torres y Victoria A. Castrillejo. Licencia Creative Commons http://www.nodosele.com/editorial

Siemens, G. (2007). *Conectivismo, una teoría de aprendizaje para la era digital*. Traducido por Diego E. Leal Fonseca https://skat.ihmc.us/rid=1J134XMRS-1ZNMYT4-13CN/George%20Siemens%20-%20Conectivismo-

Siemens, G. (2010). *Conociendo el conocimiento*. Traducción de Emilio Quintana, David Vidal, Lola Torres y Victoria A. Castrillejo. Licencia Creative Commons http://www.nodosele.com/editorial

Siemens, G. (2011). Entrevista a George Siemens desarrollador del conectivismo. Entrevista concedida a Guadalupe Vadillo. *Revista Mexicana de Bachillerato a Distancia* (6), pp. 41-47, http://bdistancia.ecoesad.org.mx

Siemens, G. & Weller, M. (2011). El impacto de las redes sociales en la enseñanza y el aprendizaje. *Revista de Universidad y Sociedad del Conocimiento* (RUSC), 8(1), 157-163 http://rusc.uoc.edu/ojs/index.php/rusc/article/view/v8n1-siemens-weller/v8n1-siemens-weller

Sobarzo-Ruíz, R., Almendras, S., Krishna, A. & Elgueta, C. (2023). Aproximación a la Teoría de la Sociedad Red, Globalización y el Conectivismo en la educación del siglo 21. *Avante*. *Revista de Humanidades y Ciencias Sociales*, 5(1), 96-105, https://doi.org/10.5281/zenodo.10440384

Spiro, R. & Jehng, J. (1990). Cognitive flexibility and hypertext: Theory and technology for the non-linear and multidimensional traversal of complex subject matter [Flexibilidad cognitiva e hipertexto: teoría y tecnología para la travesía no lineal y multidimensional de contenidos complejos]. En: D. Nix & R. Spiro (eds.), *Cognition, Education, and Multimedia* (pp. 163-205). Erlbaum

Suárez-Guerrero, C., San Martín, Á. & Limaymanta, C. (2022). Estado y diseminación del conectivismo. Análisis bibliométrico. *Education in the Knowledge Society*, (22), 1-17. https://doi.org/10.14201/eks.28212/e28212

Villalba, V. E. (2024). De las Tecnologías de la Información y la Comunicación a las Tecnologías de Aprendizaje y Conocimiento. ¿Integración o transición? *Ciencia Latina. Revista Científica Multidisciplinar,* 8(2), 5937-5954. https://doi.org/10.37811/cl_rcm.v8i2.10999

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 268-281

Ensayo/Essay

Fundamentos del conservadurismo scrutoniano

Foundations of scrutonian conservatism

Joshua Isaac Ramírez Donner

Universidad Pontificia Bolivariana, Colombia
joshuaramirezd@gmail.com
https://orcid.org/0000-0003-0610-8867

DOI https://doi.org/10.48204/2805-1815.8488

INFORMACIÓN DEL ARTÍCULO	ABSTRACT/RESUMEN

Recibido el: 16/3/2025

Aceptado el: 27/8/2025

Keywords:

Civilization, conservatism, culture, Western, sacred

Palabras clave:

Civilización, conservadurismo, cultura, Occidente, sagrado

Abstract:

This essay examines the foundations of Scrutonian conservatism, focusing on the sacredness of Western culture, which keeps its value, even without being considered a divine gift. Unlike theologically based conservatism, Scruton emphasizes secular order while acknowledging the influence of religious origins. Scruton's concept of the sacred incorporates love as a guiding principle that connects humanity through shared familial bonds, particularly within the Western context. This analysis presents Scrutonian conservatism as an attempt to offer a metaphysical framework grounded in cultural and historical content, serving as a foundation for identity, delving into his roots, his concept of the sacred, and his criteria for truth.

Resumen:

El presente ensayo analiza los fundamentos del conservadurismo de Roger Scruton, centrado en la sacralidad de la cultura occidental, que mantiene su valor, aunque ya no se considere un don divino. A diferencia de los conservadurismos con bases teológicas, Scruton enfatiza el orden secular, reconociendo la influencia de los orígenes religiosos. La concepción de lo sagrado también incluye el amor como principio rector, que conecta a la humanidad a través de vínculos familiares compartidos, especialmente en Occidente. Este análisis ofrece una perspectiva del conservadurismo scrutoniano como su intento de ofrecer un marco metafísico basado en contenidos culturales e históricos que sirve como base para la identidad, pasando por sus raíces, el concepto de lo sagrado y el criterio de verdad.

Introducción

El conservadurismo como ideología política está inmerso en una paradoja. Por un lado, la propia etiqueta de ideología resulta problemática, pues los conservadores, al enfatizar la practicidad de su pensamiento, consideran que esta implica una abstracción y tergiversación de la realidad. Por otro lado, esta misma actitud encasilla al conservador en una doctrina dictada por la conveniencia, carente de principios morales o políticos fijos y fácilmente manipulable por quienes mejor favorezcan su permanencia en el poder.

La conveniencia de identificarse con el conservadurismo genera disputas entre los propios conservadores, quienes emiten críticas como la del comentarista Ben Shapiro, quien afirma: "Trump y su gobierno no son conservadores, son pragmáticos" (Shapiro, 2016). Sin embargo, esas mismas figuras defienden su voto por él y respaldan sus decisiones, reiniciando así la paradoja mencionada. El fenómeno Trump, junto con el auge de otras formas de populismo de derecha que se cobijan bajo el conservadurismo, plantea una interrogante sobre la solidez de sus fundamentos como filosofía o ideología política.

A esto se suma la distancia que los conservadores han tomado respecto de los fundamentos basados en el naturalismo filosófico para justificar sus posturas ante una "naturaleza" en constante cambio, así como la ausencia de principios teológicos o de referencia a lo divino. El propio Michael Oakeshott, en su célebre ensayo *On Being Conservative* (1956), descartó ambos elementos como requisitos para ser conservador. Lo mismo han hecho figuras influyentes de esta corriente en el contexto latinoamericano, como Agustín Laje y Nicolás Márquez. La necesidad de ampararse en la tradición, el rechazo de la ideología, la constante adaptación a la conveniencia y la ruptura con argumentos históricamente defendidos plantean la cuestión de cuáles son, en última instancia, los fundamentos del conservadurismo.

Teniendo esto en cuenta, el filósofo conservador Roger Scruton, a menudo descartado como reaccionario o reducido a un mero recopilador del pensamiento conservador, ofrece una perspectiva matizada sobre sus fundamentos desde una óptica secular que, sin embargo, no excluye argumentos metafísicos, pues busca esclarecer la naturaleza del ser político desde el conservadurismo. Esta perspectiva merece mayor atención, no solo como una voz más dentro de esta corriente, sino como una defensa particular de sus principios, basada en un reconocimiento secular de lo sagrado. Es decir, una síntesis entre la preservación de lo empíricamente beneficiario y la necesidad de idealizar y sacralizar aspectos de nuestra cultura e historia.

Raíces en Burke y Hegel

El conservadurismo de Scruton encuentra sus bases filosóficas en su interpretación del pensamiento de Edmund Burke y Friedrich Hegel. Scruton reconoce una deuda intelectual con Burke debido a sus experiencias en Europa del Este y a las protestas de 1968 en Francia. Temía que tanto los países de Europa continental como Gran Bretaña avanzarían hacia un consenso que no solo era político y económico, sino también cultural, y que pronto sometería a Occidente a doctrinas socialistas y liberales radicales. Fue durante este período de despertar intelectual cuando Scruton encontró una profunda influencia en las obras de Edmund Burke.

> Edmund Burke argumentó en contra de la política 'geométrica', como la llamaba él, de los revolucionarios franceses, una política que proponía un objetivo racional y un procedimiento colectivo para alcanzarlo, y que movilizaba a toda la sociedad detrás del programa resultante. Burke veía la sociedad como una asociación de los muertos, los vivos y los no nacidos. Su principio vinculante no es el contrato, sino algo más parecido al amor. La sociedad es una herencia compartida por cuya causa aprendemos a circunscribir nuestras demandas, a ver nuestro propio lugar en las cosas como parte de una cadena continua de dar y recibir, y reconocer que las cosas buenas que heredamos no son nuestras para malgastarlas (Scruton, 2014, p. 27).

A lo largo de su obra, Scruton toma tres argumentos clave de Burke que sustentan su conservadurismo. En primer lugar, Burke defiende la importancia de la autoridad y la obediencia como aspectos fundamentales para el orden político (Scruton, 2017). De acuerdo con Scruton, Burke sostiene que la sociedad no se mantiene unida por derechos abstractos, como creían los revolucionarios franceses, sino por la autoridad que legitima el derecho a la obediencia, una virtud fundamental que permite que los individuos sean gobernados (Burke, 1790/1987). En ausencia de tal autoridad, las sociedades corren el riesgo de fragmentarse en una colección caótica de individuos aislados.

En segundo lugar, Burke defiende el valor de la tradición, el prejuicio y la costumbre frente a los planes de reforma inspirados en la Ilustración (Scruton, 2017). La tradición, para él, no es una reliquia estática, sino un proceso dinámico que adapta continuamente el pasado al presente y el presente al pasado. Scruton enfatiza que esta visión rescata el estudio de la historia de la mera abstracción académica y reafirma la importancia de la continuidad cultural, particularmente a través del arte, la literatura y la filosofía. El respeto por la costumbre, sugiere Burke, es una virtud, no un signo de complacencia, como creían muchos de sus contemporáneos.

En tercer lugar, Burke (1790/1987) critica la teoría del contrato social, particularmente el contrato defendido por Rousseau. Si bien la sociedad puede entenderse como un contrato, Burke argumenta que también es un fideicomiso entre los vivos, los muertos y los no nacidos. El rechazo de los revolucionarios a los derechos ancestrales no solo malgastó los recursos heredados, sino que también desheredó a las futuras generaciones (Scruton, 2017). La visión de Burke sobre la sociedad no está, por lo tanto, enraizada en un contrato, sino en la responsabilidad y el deber que los vivos tienen de mejorar y preservar la herencia que han recibido y que deberán transmitir intacta a generaciones del futuro.

A través de Burke, Scruton (1980, 2007, 2014, 2017) apela a la autoridad, la tradición y la reformulación del contrato social para preservar la costumbre, las instituciones sociales y las relaciones humanas de carne y hueso, ya forjadas por lazos de confianza y responsabilidad que desvelan una convivencia empíricamente defendible. Pero es a través de su interpretación de Hegel que Scruton sienta las bases para un argumento metafísico en favor del conservadurismo. Scruton se inspira en Hegel para defender el sentido de pertenencia, el cual dota de protección a valores como la libertad y la igualdad. Para Scruton, al igual que para Hegel (1820/1991), la realidad no es simplemente la suma de objetos aislados o hechos individuales, sino un proceso interconectado y autoorganizado. En el caso de Hegel, la realidad no se restringe a la mente del sujeto individual, sino que es un proceso objetivo, universal y social. En este contexto, el idealismo de Hegel está profundamente ligado a la historia y a la cultura, donde la razón se presenta como un proceso autodeterminante.

Para Scruton, la libertad se alinea estrechamente con la definición de Hegel, en contraste con la de Immanuel Kant. Es la realización de la autodeterminación, en la que los individuos no solo toman decisiones personales, sino que también descubren su verdadera voluntad y racionalidad dentro de las instituciones sociales y la vida ética de un Estado racional, armonizando así la autonomía personal con los principios universales. La libertad es más que la ausencia de restricciones externas al movimiento; para ser verdaderamente libres, es necesario estar en armonía con el entorno social, histórico y ético propio (Scruton, 1980). Esta comprensión presenta una afirmación metafísica sobre la naturaleza de la libertad como un concepto fundamental que sustenta la estructura del mundo social, moldeando las relaciones sociales, la existencia humana y la agencia individual. En esencia, se trata de estar en relación con objetos o entidades que son significativos para el carácter.

No obstante, para Hegel, la fuerza motriz de la historia radica en la actualización progresiva de ambas formas de libertad. Scruton, especialmente en sus escritos más recientes, comparte esta visión, considerando que la libertad es fundamental para una coordinación social efectiva. La verdadera libertad no se alcanza aislándose del mundo, sino participando en un orden social que refleje los valores éticos más profundos y el potencial humano de cada individuo (Scruton, 1980). En este sentido, la libertad se realiza dentro del contexto de relaciones comunitarias y valores compartidos. Como sugiere Scruton, uno debe pertenecer para ser, es decir, sentirse en casa consigo mismo a través de sus conexiones con los demás.

Este aspecto dual de la libertad está estrechamente vinculado con el concepto de alienación de Hegel, una noción que Scruton reconoció como profundamente significativa, particularmente en lo que respecta a la alienación subjetiva. Esta alienación subjetiva surge cuando los individuos no logran sentirse en casa en el mundo que habitan, incluso cuando los arreglos sociales, en teoría, respaldan su autorrealización. (Hegel, 1820/1991) Para superar esta alienación, Hegel sostiene que los individuos deben reconocer que, aunque el mundo que los rodea pueda parecer extraño o incluso opresivo, en realidad ofrece las condiciones necesarias para su libertad y florecimiento. Cuando las personas comprenden que el mundo social e histórico está estructurado para facilitar su desarrollo, comienzan a sentirse en casa dentro de él, y su sentido de alienación se disuelve. Para Hegel, este proceso de reconciliación es esencial para superar la alienación subjetiva, y es precisamente lo que Scruton identifica como la fuente que le otorga sentido a la sociedad secular: un sentido de "retorno a casa", es decir, un sentido de pertenencia.

La base de su pensamiento conservador está vinculado al deseo de preservar lo que amamos, un deseo profundamente arraigado en un sentido de pertenencia (Scruton, 1982). De esta manera, su visión resuena con la exploración de Hegel sobre la conexión entre la libertad y el amor. Para Hegel, el amor no es simplemente una emoción, sino una expresión de nuestra verdadera humanidad. El amor representa una conexión ética entre los individuos, y la capacidad de ser libre está estrechamente vinculada con la capacidad de amar (Hegel, 1820/1991). Aunque Scruton nunca compartió la visión de que la historia está gobernada por un plan de desarrollo en el que cada etapa histórica representa una encarnación más adecuada de la razón, definida por las acciones, pasiones y ambiciones de ciertos individuos para cumplir con su propósito, sí considera que el legado perdurable de determinadas etapas históricas constituye la base de nuestra herencia (Scruton, 2014). Esta herencia, a su vez, dio origen a un nivel de superioridad entre los pueblos de habla inglesa, una superioridad que, según él, ha sido atacada por lo que denomina una "cultura de repudio".

En el centro del conservadurismo de Scruton se encuentra la creencia de que la sociedad precede al individuo. En esta visión, inspirada en Hegel, el individuo es moldeado por las condiciones históricas y sociales particulares que hereda, incluyendo los valores, costumbres y expectativas de su comunidad. Estos valores no son arbitrarios ni subjetivos; más bien, forman un orden social coherente en el que el individuo participa y en el que, hasta cierto punto, se convierte. Desde la perspectiva conservadora, estos valores y costumbres heredados merecen respeto, y cualquier rechazo de estos debe ocurrir dentro del contexto de ese orden social. Rechazar dichos valores es romper con el mismo tejido de la sociedad que hace que ese rechazo tenga sentido. Cuando estos valores pierden su autoridad, requieren de una renovación de las tradiciones o la introducción de una forma de justicia natural que les otorgue universalidad.

Esto es parte de lo que Scruton encuentra objetable en el liberalismo a un nivel más fundamental: su dependencia del concepto del individuo autónomo, quien se presume que está desprovisto de los lazos sociales y las obligaciones que hacen posible un orden social estable (Scruton, 1980). Scruton argumenta que conceptos como la libertad y los derechos no pueden existir en un vacío; están inextricablemente ligados a tradiciones particulares y contextos culturales. La noción liberal de individuos abstractos, unidos solo por un contrato social, con sus deberes sociales derivados únicamente de elecciones autónomas, es inaceptable para los conservadores. Según la visión de Scruton, los lazos de la sociedad trascienden la elección individual, y es imposible derivar los fines de la conducta humana únicamente de la autonomía individual.

La autoconciencia y la libertad se desarrollan a través de las relaciones con los demás. La libertad humana, para Scruton, evoluciona superando el conflicto y reconociendo los derechos y deberes mutuos, lo que finalmente fomenta un sentido tanto de valor personal como colectivo. Para los conservadores, las instituciones de la ley, la educación y la política no son estructuras abstractas impuestas desde arriba, sino que son fundamentales en el proceso de convertirse en individuos plenamente libres y autoconscientes (Scruton, 2014). Estas instituciones ayudan a los individuos a vivir como agentes responsables dentro de la sociedad, conscientes de sus obligaciones y roles.

La practicidad de lo sagrado

El *oikos*, como lo menciona Scruton (2014), forma la base de la estructura ética que compartimos con los demás y que consideramos buena en sí misma y, por lo tanto, digna de ser preservada. Pero esta visión y aceptación del bien, o lo bueno, implica una afirmación metafísica que, para Scruton, no está simplemente arraigada en lo que es efectivo o útil, sino en lo que es verdaderamente beneficioso para nosotros, lo que ha contribuido a nuestro carácter, desarrollo, bienestar e identidad (Scruton, 2014, p. 28). Esto es lo que hace que estas cosas sean buenas e infunde en nosotros un sentido de deber para preservarlas para las generaciones futuras, tal como las hemos heredado de aquellos que vinieron antes que nosotros.

Para Scruton, siempre es correcto conservar lo que valoramos, especialmente cuando se nos proponen alternativas que son peores. Esta ley a priori de la razón práctica es la verdad fundamental que sustenta su conservadurismo. Lo que apreciamos en la sociedad incluye, en su núcleo, conceptos como el sacrificio, el honor militar, el apego familiar, las estructuras y los contenidos educativos, las instituciones benéficas y las normas de buena conducta.

En este sentido, el valor de la tradición, según Scruton, no radica en un conocimiento teórico de hechos y verdades, ni se limita a un saber común. Se trata más bien de un dominio práctico en la interacción social: saber cómo llevar a cabo una tarea con éxito, donde el éxito no se mide por un objetivo exacto o predefinido, sino por la armonía del resultado con nuestras necesidades e intereses humanos. Saber cómo comportarse en sociedad, qué decir y qué sentir son cosas que adquirimos a través de la inmersión social. No pueden enseñarse únicamente con instrucciones explícitas, sino que se transmiten por ósmosis. Sin embargo, a quien no ha adquirido estas habilidades se le considera, con razón, ignorante.

La cultura desempeña el papel de dotar de identidad a los miembros de la sociedad civil. En su texto *Culture Counts* (2007), Scruton equipara la cultura con el quehacer de las élites. La defensa de la alta cultura es crucial, ya que ha sido rechazada no sólo por los radicales, sino incluso por los liberales moderados, quienes consideran que su importancia se reduce a la tecnología, la ciencia y todo lo que imite el progreso. Para Scruton, la cultura es una adquisición que abre tanto las mentes como los corazones, creada por élites inmersas en el patrimonio intelectual y artístico, pero que resuena con emociones universalmente humanas. Sirve como un depósito de conocimiento moral y como una visión compartida de la dignidad.

Mientras que la ciencia y la tecnología ofrecen progreso, no pueden reemplazar los fundamentos morales que proporciona la cultura. La cultura es vital para el sentido de pertenencia, requiriendo estudio activo, renovación y transmisión para evitar el declive intelectual y moral. Los juicios estéticos a los que se someten las obras de arte y literatura determinan guías para quienes las heredan y las apropian.

Aunque la cultura pueda surgir de círculos elitistas, su significado resuena ampliamente con emociones y aspiraciones humanas. Si bien la civilización es una entidad social más amplia, caracterizada por la uniformidad religiosa, política, legal y consuetudinaria a lo largo de un período extenso de tiempo, proporciona a sus miembros conocimientos socialmente acumulados y estructuras institucionales (Scruton, 2007). Las civilizaciones pueden contener múltiples tradiciones culturales y existir simultáneamente o de manera sucesiva, pero la cultura es el medio mediante el cual la civilización toma conciencia de sí misma y define su visión del mundo.

Para Scruton, el respeto por la tradición es una virtud, no un signo de complacencia. Es un proceso dinámico que adapta continuamente el pasado al presente y viceversa. Esta visión rescata el estudio de la historia de la mera abstracción académica y reafirma la importancia de la continuidad cultural (Scruton, 2017, p. 45). Por ejemplo, nuestra comprensión de la felicidad se encuentra en el sacrificio promovido por el cristianismo, que enfatiza el perdón a través de la renuncia a la venganza y la aceptación de los demás tal como son. También aboga por el concepto de ley como un medio para resolver conflictos, tratando a cada parte como un agente responsable. En este sentido, aunque el logro occidental de la ciudadanía es, en efecto, una relación entre extraños cuyo significado se limita a la esfera privada, solo podrá perdurar si está vinculado a un sentido de identidad basado en tradiciones expresadas en grandes obras de arte, literatura, filosofía y derecho; recursos que nutren espiritualmente y que serán heredados por las futuras generaciones.

Lo que inicialmente puede parecer abstracto, en realidad se manifiesta como algo concreto y práctico a través de una cualidad que obliga a los individuos a seguir ciertos principios. Estos principios se reflejan en la seguridad de una ley imparcial, la protección del entorno como un bien común, la cultura abierta y cuestionadora promovida por la educación, y los procedimientos democráticos. Esto contrasta con la teoría del contrato social defendida por muchos liberales, la cual, al no reconocer la dependencia mutua previa ni la cohesión forjada por el territorio y la jurisdicción, se mantiene como un mero experimento mental. Carece de las herramientas necesarias para orientar a los individuos sobre cómo podrían ser gobernados. Además, los miembros del contrato ya deben pertenecer a una comunidad, pues ya se han concebido como tal a través del largo proceso de interacción social. Para Scruton, la adhesión a la visión de Burke sobre la sociedad –como una asociación entre los muertos, los vivos y los no nacido– no se basa en un principio contractual, sino en algo más cercano al amor y la confianza. En palabras de Scruton:

> Debemos invertir nuestro amor y deseo en cosas a las que asignamos un valor intrínseco, en lugar de un valor instrumental, para que la búsqueda de medios

> pueda llegar a descansar, para nosotros, en un lugar de fines. Eso es lo que queremos decir con asentamiento: poner el oikos de vuelta en la oikonomía. Y eso es de lo que trata el conservadurismo (2014, p. 32).

En el conservadurismo, la virtud se convierte, finalmente, en un principio político. En este sentido, la base metafísica descrita antes se expresa más claramente a través del individuo y su rol en la sociedad civil, que se forma dentro del marco sagrado de una estructura histórica y cultural realizada mediante un ejercicio pragmático. Los conservadores reconocen que algunos problemas pueden no tener soluciones políticas inmediatas y, por lo tanto, confían en el marco establecido por las instituciones heredadas para ayudar a resolverlos (Scruton, 1980). La necesidad de ser prácticos requiere esta base metafísica, arraigada en una visión idealizada de la historia y la cultura compartida; esencialmente, preguntando: ¿qué es lo que deseamos preservar?

Criterios de verdad

A pesar de apelar a Burke para defender el valor de la tradición, el prejuicio y las costumbres frente a las reformas progresistas inspiradas en la Ilustración, Scruton también recurre a este legado histórico por el papel fundamental que le asignó al uso de la razón (Scruton, 2017). Este es el legado que, según él, nos permite comprometernos de manera razonable y definir el papel de la razón, especialmente cuando nos enfrentamos al rechazo de la verdad objetiva, como ocurre en el pragmatismo de Richard Rorty, el cual, según Scruton, podría llevarnos a una peligrosa deriva hacia el relativismo moral y el nihilismo intelectual.

Dicho de otra forma, aunque Scruton reconoce la necesidad del pragmatismo como un eje fundamental en la política y en la encarnación de los valores impartidos por la cultura, también defiende la objetividad y la verdad universal, en contraste con los valores de pragmatistas como Rorty, quienes consideraban que la verdad dependía únicamente de un consenso general (véase Scruton, 2014). Scruton se esfuerza por definir la verdad como objetiva e independiente del acuerdo humano, descubriéndose a través de la indagación racional y fundamentada en una correspondencia con la realidad que habitamos. Para él, la verdad no es una construcción social ni un producto de acuerdos intersubjetivos, y este razonamiento se inserta dentro de una tradición intelectual arraigada en los ideales de la Ilustración, donde la razón universal y el juicio moral juegan un papel esencial. Para Scruton, la búsqueda de la verdad es fundamental para el florecimiento humano y la salud moral de la sociedad.

La crítica de Scruton al multiculturalismo revela su creencia en verdades morales objetivas que aplican a todos los seres humanos, independientemente de las creencias personales. Estas verdades, sostiene, están fundamentadas en la razón, la tradición de la Ilustración occidental y la herencia cultural del Estado-nación, junto con las formas de vida social que han surgido dentro de él (Scruton, 2014). Por eso, para Scruton, la creación de una nación basada en la verdad conservadora de la lealtad nacional se logra mejor dentro de un contexto occidental, donde las diferencias de creencias pueden coexistir y dar lugar a una identidad compartida. La lealtad nacional, en su opinión, trasciende las divisiones de familia, religión o credo, fomentando una obligación recíproca entre los ciudadanos, quienes deben lealtad los unos a los otros en base a su nacionalidad común. Este vínculo hace posibles las libertades –como la adoración, la libertad de expresión y de conciencia– no solo como viables, sino también como compatibles con un sentido colectivo de unidad. El Estado-nación, a su vez, es responsable ante todos sus ciudadanos, derivando su legitimidad de esta lealtad compartida, que define sus fronteras territoriales y limita su autoridad.

La admiración de Scruton por el derecho común inglés subraya aún más su convicción de que la verdad emerge a través del juicio sobre la coexistencia, entrelazando derechos y deberes en un proceso de descubrimiento. Nuestras capacidades racionales, argumenta, están moldeadas por la configuración de nuestra identidad cultural e histórica, preservada por un amor moralmente superior al odio. El derecho común, no es una imposición desde arriba, sino un sistema que crece orgánicamente a partir de la libre asociación de individuos. Este énfasis en un orden espontáneo y descentralizado se alinea con la visión más amplia de Scruton de la sociedad como una red compleja de costumbres, hábitos y normas informales que regulan el comportamiento y fomentan la cohesión. Para Scruton, la sociedad no es meramente una colección de reglas formales; es una red viva y en evolución de prácticas compartidas que emergen de las experiencias y elecciones de los individuos dentro de la comunidad, guiadas por principios que facilitan la convivencia pacífica.

De la misma manera, Scruton sostiene que el individuo tiene un deseo de vivir en comunidad y resalta la importancia de preservar las instituciones que encarnan los valores y costumbres de la sociedad civil. Los conservadores se oponen a la creación de instituciones basadas en teorías políticas abstractas, prefiriendo en su lugar aquellas que surgen de manera orgánica de las necesidades y tradiciones de la nación. Aunque el Estado desempeña un papel importante, no es la fuerza central en el conservadurismo; más bien, sirve como un medio para mantener la soberanía, proporcionar autoridad ceremonial y reconciliar los intereses sociales existentes. Según Scruton, los conservadores reconocen la sabiduría incrustada en las tradiciones de largo plazo y son reacios a apoyar reformas radicales que puedan perturbar el equilibrio y la estabilidad que estas instituciones brindan.

El conservadurismo scrutoniano pretende defender la oportunidad de vivir nuestras vidas cómo elegimos, la seguridad de la ley imparcial, la protección de nuestro entorno como un bien compartido, la cultura abierta y de investigación fomentada por la educación, y los procedimientos democráticos que nos permiten elegir a nuestros representantes y aprobar nuestras propias leyes. Estos son solo algunos de los muchos ejemplos de instituciones y costumbres familiares del mundo de habla inglesa, que Scruton vio como cada vez más amenazadas a lo largo de su carrera. La práctica política conservadora scrutoniana es pragmática y local, prefiriendo el cambio incremental sobre las transformaciones radicales. Reconoce que algunos problemas pueden no tener soluciones inmediatas, pero sugiere que las instituciones heredadas, fundamentadas en el Estado de derecho, ofrecen los medios más confiables para resolverlos.

Conclusión

En conclusión, los fundamentos del conservadurismo scrutoniano se basan en una metafísica que dota de sacralidad a la cultura occidental, cuya aura sigue siendo valiosa, aunque ya no se considere un don divino. Esto no implica que el conservadurismo sea completamente secular; más bien, reconoce que gran parte de lo que valoramos está profundamente marcado por sus orígenes religiosos. Sin embargo, Scruton se diferencia de la visión de que el conservadurismo se construye sobre fundamentos teológicos, en los cuales el orden democrático occidental se considera divinamente ordenado y las asociaciones humanas están guiadas por el Espíritu Santo (Scruton, 2014). Para Scruton, los conservadores británicos suelen ser menos propensos a aceptar tales visiones teológicas, debido a la experiencia del Imperio y a la necesidad de mantener el orden civil entre pueblos que no compartían una cosmovisión cristiana.

Scruton sostiene una concepción de lo sagrado que permite que conceptos como la ley natural y universal sigan siendo relevantes dentro de las instituciones que hemos heredado, valorado y justificado como verdaderas. Aunque la religión pueda servir como raíz de las comunidades y ofrecer consuelo a los individuos, juega un papel más ceremonial en la vida del Estado, el cual está fundado en principios seculares—en particular, el principio de la libertad religiosa. Para Scruton, el ámbito de los valores religiosos está abierto para todos: los individuos pueden unirse a iglesias y templos, explorar la santidad y la rectitud, y experimentar la paz, la esperanza y el consuelo que la religión ofrece. Sin embargo, también deben reconocer y respetar el derecho de los demás a ser diferentes.

La defensa del amor como principio y del deseo de preservar está en el corazón del conservadurismo scrutoniano, una cualidad intrínseca que conecta a todos los seres humanos a través de un sentido compartido de lo familiar. Este amor se manifiesta de manera más profunda en Occidente, donde sus costumbres e instituciones modelan las relaciones humanas, y fomentan su propio crecimiento. A partir de este reconocimiento de valores e ideales, Scruton extrae las justificaciones necesarias para los argumentos políticamente activos en cuestiones como el ambientalismo.

El proyecto de Scruton es una filosofía de la constitución moral y cultural del hombre. Scruton toma el análisis sociohistórico de Hegel y lo convierte en una dinámica de construcción de la identidad y el alma. En este sentido, los individuos reconocen libremente sus obligaciones y se apropian de una vida amparada en el deber y la piedad. Para el conservador, el hombre es libre y sumiso, dependiente de la comunidad, pero capaz de reconocer y acoger autónomamente los estándares morales que de ella emanan.

El conservadurismo de Scruton fundamenta el conservatismo en la creencia de que los ideales y las prácticas que heredamos poseen un valor sagrado intrínseco y merecen ser venerados, expresándose a través del amor como principio rector y orientado a alcanzar fines que sirvan a propósitos más elevados. A diferencia del liberalismo y el socialismo, que a menudo son criticados por enfocarse radicalmente en principios abstractos que carecen de sustancia práctica o aplicabilidad, el conservadurismo scrutoniano proporciona una base metafísica fundamentada en contenidos culturales, históricos y morales que sirven como base para la identidad. Quiénes somos está entrelazado con las instituciones y principios en los que creemos.

Referencias bibliográficas

Oakeshott, M. (1956). On being conservative [Sobre ser conservador]. http://www.geocities.com/Heartland/4887/conservative.html

Burke, E. (1987). *Reflections on the revolution in France* [Reflexiones sobre la revolución Francesa] (J. G. A. Pocock, Ed.). Penguin Classics. (Obra original publicada en 1790)

Hegel, G. W. F. (1991). *Elements of the philosophy of right* [Elementos de filosofía del derecho] (A. W. Wood, Ed. & H. B. Nisbet, Trans.) [Kindle edition]. Cambridge University Press. (Obra original publicada en 1820)

Scruton, R. (1980). *The meaning of conservatism* [El significado del conservadurismo]. St. Agustine press.

Scruton, R. (1982). *A dictionary of political thought* [Un diccionario de pensamiento político]. Pan Reference.

Scruton, R. (2007). *Culture counts: Faith and feeling in a world besieged* [La cultura cuenta: fe y sentimiento en un mundo asediado]. Encounter Books.

Scruton, R. (2014). *How to be a conservative* [Cómo ser un conservador]. Bloomsbury Continuum

Scruton, R. (2017). *Conservatism: An invitation to the great tradition* [Conservadurismo: una invitación a la gran tradición]. Points Book

Shapiro, B. (2016, November 23). Is Donald Trump a pragmatist? [¿Es Donald Trump un pragmatista?] *National Review*. https://www.nationalreview.com/2016/11/donald-trump-pragmatist-not-conservative/

Analítica (5), Oct. 2025 – Sept. 2026
ISSN – L 2805 – 1815, pp. 282-293

Ensayo/Essay

Comentarios sobre la crítica de Schopenhauer a Kant con relación a la cosa en sí y la ley de causalidad

Comments on Schopenhauer's critique of Kant regarding the thing-in-itself and the law of causality

Ruling Barragán Yáñez

Universidad de Panamá, Panamá
ruling.barragan@up.ac.pa
https://orcid.org/0000-0002-3237-5726

DOI https://doi.org/10.48204/2805-1815.8489

INFORMACIÓN DEL ARTÍCULO

Recibido el: 27/7/2025

Aceptado el: 26/9/2025

Keywords:

Kant, Schopenhauer, thing-in-itself, causality, idealism.

Palabras clave:

Kant, Schopenhauer, cosa en sí, causalidad, Idealismo

ABSTRACT/RESUMEN

Abstract:

The purpose of this text is to comment on Schopenhauer's critique of Kantian thought, particularly the idea of the thing-in-itself and the notion of causality as upheld by the philosopher of Königsberg. First, I will address Schopenhauer's arguments against Kant concerning the thing-in-itself. Next, I will consider Schopenhauer's line of reasoning against Kant's view of causality. Finally, in the absence of a strict conclusion, I will share a few conjectures that might shed light on the argumentative deficiencies that Schopenhauer attributes to his master.

Resumen:

El propósito de este texto es comentar la crítica que hace Schopenhauer al pensamiento de Kant, en particular, a la idea de la cosa en sí y la noción de causalidad sustentadas por el filósofo de Königsberg. Así pues, en primer lugar, abordaré los argumentos de Schopenhauer contra Kant acerca de la cosa en sí. Seguidamente, trataré la argumentación schopenhaueriana contra Kant acerca de la causalidad. Por último, a falta de una conclusión en sentido estricto, compartiré unas conjeturas que podrían explicar las faltas argumentativas que Schopenhauer le reprocha a su maestro.

La cosa en sí

Según Schopenhauer, el principal mérito de Kant lo constituye el haber distinguido el fenómeno de la cosa en sí, al declarar que el mundo visible no es sino fenómeno y las leyes que le son inherentes no tienen aplicabilidad válida fuera de él. Sin embargo, resulta curioso que –señala Schopenhauer– Kant no encuentra veracidad en una 'simple e innegable verdad' que yace al pie de aquella declaración: a saber, "que no hay objeto sin sujeto". De esta última proposición se deduce que el objeto, debido a que sólo existe en relación con el sujeto, es dependiente y condicionado por él. Así, el mundo del fenómeno no existe ni puede existir en sí mismo e incondicionalmente.

Debido a lo anterior, Schopenhauer considera que Kant no le ha hecho justicia a Berkeley, pues éste último ya había establecido que el mundo fenoménico depende enteramente de un sujeto que le condiciona. No obstante, Berkeley 'no alcanzó a deducir las adecuadas y pertinentes conclusiones de tal afirmación, por lo cual fue parcialmente malentendido y no se le prestó la debida atención'. Schopenhauer señala que la posición de Kant con relación a la consideración de Berkeley resulta de 'un visible temor a un decisivo idealismo'; Schopenhauer mismo encuentra pasajes en la primera *Crítica de la Razón Pura* que son una directa contradicción con lo que posteriormente sostuvo Kant. En esta primera edición, algunos importantes pasajes manifiestan una posición idealista en el sentido de Berkeley. Así, Schopenhauer encuentra 'con gran gozo' que, en esta primera edición, "aunque Kant no utiliza la fórmula 'no hay objeto sin sujeto', sin embargo, declara con tanto énfasis, como hace Berkeley y yo, que el mundo exterior que yace ante nosotros en el tiempo y en el espacio es una mera representación del sujeto que lo conoce" (Schopenhauer, 1958, p. 434-435). No obstante, los pasajes en que Kant expone decisivamente su idealismo son suprimidos en la segunda edición de la *Crítica*.

En una carta fechada el 24 de agosto de 1837[1] Schopenhauer discutió detalladamente con Rosenkranz sobre estas supresiones y las conjeturas que el mismo Schopenhauer había desarrollado para explicar la mutilación que Kant hace de su propia obra. "El principal pasaje de esta carta fue incluida por Rosenkranz en su prefacio al segundo volumen de las obras completas de Kant editadas por él…" (Schopenhauer, 1985, p. 435). Debido a la insistencia de Schopenhauer, Rosenkranz restauró la *Crítica* a su forma original. "De esta forma él [refiriéndose a Rosenkranz] otorgó un inestimable servicio a la filosofía; ... [y] por esto debemos estarle siempre agradecidos" (p. 435). A juicio de Schopenhauer nadie puede ni siquiera imaginar que ha comprendido la enseñanza kantiana, si no ha leído la primera edición del trabajo de Kant. "La segunda y subsiguientes ediciones de la *Crítica* son un texto mutilado, echado a perder, y hasta cierto punto espurio" (p. 435).

Una de las principales inconsecuencias de Kant al no admitir la posición de Berkeley está en la doctrina de la cosa en sí. La manera en que Kant introduce la noción de la cosa en sí, afirma Schopenhauer, se halla en contraposición con el punto de vista idealista. Esta introducción trajo consigo inconsistencias irremediables. Ya G. E. Schulze se había encargado de demostrar la invalidez de la introducción de la cosa en sí perpetrada por Kant. Secundando esta crítica, Schopenhauer señala que:

> Kant basa la suposición de la cosa en sí ... en una conclusión de acuerdo a la ley de la causalidad, a saber, que la percepción empírica, o más correctamente la sensación de nuestros órganos de los sentidos de la cual procede, debe tener una causa externa. Ahora, de acuerdo a su correcto descubrimiento, la ley de la causalidad nos es descubierta a priori, y consecuentemente es una función de nuestro intelecto; por lo tanto, es de origen subjetivo. Más aún, la sensación misma, a la cual aplicamos aquí la ley de la causalidad, es innegablemente subjetiva; y finalmente, aún el espacio, en el cual, por medio de tal aplicación, colocamos la causa de la sensación como objeto, es una forma de nuestro intelecto dada a priori y es consecuentemente subjetiva. Luego, toda percepción empírica resulta enteramente sobre un fundamento subjetivo, como una mera ocurrencia en nosotros, y nada completamente diferente e independientemente de ésta puede ser traída como cosa en sí o mostrada como una suposición necesaria. La percepción empírica es y sigue siendo nuestra representación; es el mundo como representación (Schopenhauer, 1958, p. 436).

De este modo, resulta inaceptable que la admisión de la cosa en sí resulte de la aplicación de la ley de la causalidad, pues esta es de origen subjetivo, y por lo tanto permanece en ese ámbito, junto con las formas a priori de la sensibilidad, tiempo y espacio. De esto se sigue que no puede haber una relación causal entre cosa en sí y la representación o percepción empírica, pues la cosa en sí es noumenon por definición y la categoría de causalidad es sólo aplicable a la esfera del fenómeno[2].

Según Schopenhauer, gran parte del anterior error se debe también a que Kant no ha distinguido adecuadamente el conocimiento que proviene de la percepción de aquel cuyo origen reside en el conocimiento abstracto. Los 'objetos de la experiencia' de los cuales habla Kant no son tratados –a juicio de Schopenhauer– ni como percepciones ni tampoco como conceptos abstractos, sino como algo diferente de ambos, siendo, sin embargo, ambos al mismo tiempo, 'lo cual es una expresa absurdidad e imposibilidad' (1958, p. 437). De acuerdo con Schopenhauer, Kant no entiende por falta de una comprensión adecuada, si su 'objeto de experiencia', es decir, aquel que surge en función de la aplicación de las categorías, es una representación (o percepción empírica), o simplemente, un concepto abstracto. 'Aunque resulte en demasía extraño e increíble –dice Schopenhauer– en la mente de Kant existe algo entre las dos alternativas'. La noción de 'objeto de experiencia' en Kant se basa en un análisis que no esclarece con suficiente claridad a qué clase de objetos se está refiriendo. Kant parece confundir en su 'objeto de experiencia' lo que es representativo con lo que es conceptual.

A juicio de Schopenhauer, de aquella falta de distinción entre lo empírico y lo abstracto se debe el carácter abstruso de la Lógica Transcendental. En ella, Kant establece que nuestro conocimiento proviene de dos fuentes: (1) la capacidad de recibir representaciones y (2) la capacidad de conocer un objeto a través de estas representaciones. "Con la primera el objeto nos es dado; con la segunda, es pensado". Esto es inaceptable para Schopenhauer, ya que, según lo anterior, la impresión sería una representación; de hecho, un objeto (1958, p. 438). Pero la impresión, considerada en sí, no es más que una mera sensación, una modificación de los órganos de los sentidos, según el análisis de Schopenhauer en *De la Cuádruple Raíz*. Esta modificación se convierte en representación, sólo a través de la ley de causalidad. Sólo así la representación es idéntica al objeto. El conocimiento de percepción o la representación está, luego de aplicación de la ley causal, completamente acabada; no se necesitan, por tanto, ni conceptos o pensamiento abstracto. Si al pensar abstracto se añade la representación, entonces se abandona a esta última, entrándose así a una nueva clase de representaciones, los conceptos, que no son ni intuibles o perceptibles empíricamente.

Al introducir 'la actividad de los conceptos' en la percepción o representación empírica, Kant incurre en aquella confusión que Schopenhauer condena severamente. Kant establece que la percepción sólo deviene en objeto a través del pensamiento. Sin embargo, al establecer aquel supuesto hecho, el 'objeto del pensamiento', pasaría a ser algo individual, un objeto real. Pero de esta forma, indica Schopenhauer, el pensamiento pierde su esencial de universalidad y abstracción, y en vez de conceptos universales recibe como su objeto cosas individuales (1958, p. 439). Esto es, obviamente, algo inadmisible y se debe a la confusión de Kant, por no tener una 'clara y distinta' noción de las representaciones de percepción y las representaciones abstractas. Por ello, Kant tiende, implícitamente, a tratar algo así como un cruce entre las dos clases de representaciones, cuando describe al objeto del entendimiento que producen las categorías, como un objeto de experiencia. "Es difícil creer- señala Schopenhauer- que, en el caso de este objeto del entendimiento, Kant se imaginó a sí mismo algo bastante definido y realmente distinto[3].

De acuerdo con Schopenhauer, el 'objeto de las categorías' del cual habla Kant no es la cosa en sí, sino un extraño híbrido. "Es el ‘objeto en sí’, un objeto que no requiere sujeto, una cosa individual, pero que no está en el tiempo y en el espacio debido a que no es perceptible; es el objeto del pensamiento, aunque no es un concepto abstracto" (1958, p. 444). Según este enfoque, entonces, Kant efectúa una distinción triple: (1) la representación; (2) el objeto de la representación; (3) la cosa en sí. No obstante, una distinción de tal índole no es posible, si hemos comprendido bien a Berkeley, y si comprendemos consecuentemente a Kant, sólo podremos admitir dos cosas: la representación y la cosa en sí. Nada más.

Las críticas contra la causalidad

Schopenhauer está en total desacuerdo con la prueba ofrecida por Kant en favor de la aprioridad de la ley de causa y efecto, o principio de causalidad. Según Schopenhauer, ésta se encuentra básicamente expuesta en el siguiente pasaje de la *Crítica* según cita en su tesis doctoral:

> La síntesis de lo diverso, necesaria en todo conocimiento empírico, operada por medio de la imaginación, da la sucesión, pero una sucesión aún no determinada: es decir, deja indeterminado cuál de los dos estados percibidos precede al otro,

> no sólo en mi imaginación, sino en el objeto. El orden determinado de esta sucesión, por el cual la percepción se convierte en experiencia, es decir, justifica juicios objetivamente válidos, sólo penetra mediante el concepto puro intelectual de causa y efecto. De este modo, el principio fundamental de la relación causal es la condición de posibilidad de la experiencia, y como tal nos es dada a priori (Schopenhauer, 1989, p. 133).

Según este pasaje, señala Schopenhauer, el orden de la sucesión de los cambios de los objetos sólo nos es conocido como objetivo por medio de la causalidad de estos; “Kant afirma… que la objetividad de la sucesión de representaciones… nos es conocida solamente por la regla mediante la cual se suceden unos a otros, esto es, por la ley de causalidad…". (1989, p. 134). Kant asegura que, si fuese por la mera percepción, la relación objetiva de los fenómenos que son sucesivos quedaría indeterminada, ya que esta percepción es subjetiva y, por lo tanto, no puede validar la objetividad de la sucesión, a menos que se apoye en la ley de la causalidad. Así, Kant argumenta que uno podría invertir a voluntad el orden de las sucesiones; para ello, toma como ejemplo la percepción de una casa cuyas partes uno podría considerar en cualquier orden de sucesión, v.gr., de arriba hacia abajo o de abajo hacia arriba, por lo cual la determinación objetiva de la sucesión de las percepciones estaría subjetivamente fundada (1989, p.135). Por otro lado, Kant nos ofrece también un ejemplo en el cual la sucesión no es determinada subjetivamente: es el caso de la percepción de un barco descendiendo por un río; de acuerdo con éste, siempre percibiríamos al barco descendiendo desde arriba hacia abajo. En este caso, la sucesión de las posiciones del barco es percibida, en todo momento, de arriba hacia abajo, sin que uno las pueda hacer variar a voluntad.

En contra de este argumento, Schopenhauer asevera que, en ambos ejemplos, la sucesión es objetiva. Según nuestro filósofo, lo que pasa es que, en el ejemplo de la casa, Kant no toma en cuenta el cuerpo del observador, es decir, el ojo. Tanto en el caso de la casa como el del barco, la sucesión de las percepciones es objetiva, en cuanto que ambos son las mutaciones de posición de dos cuerpos entre sí" (p. 135). Así, nos elucida Schopenhauer:

> En el [ejemplo de la casa] ... uno de [estos cuerpos]… ... es el propio cuerpo del observador, y, por cierto, sólo una parte del mismo, el ojo, y el otro, la casa,

> respecto de cuyas partes la posición del ojo varía sucesivamente. En el [ejemplo del barco], varía el barco su posición con respecto al río; por tanto, también se trata de una mutación de posición entre dos cuerpos ... [La importante diferencia, que Kant no toma en cuenta, es que en el caso de la casa] ... la mutación procede del propio cuerpo del observador… que no por esto deja de ser un objeto entre los demás objetos, y, por tanto, está sometido a las leyes de este mundo objetivo corporal. [En la percepción de las partes de la casa], [e]l movimiento […] [del] cuerpo según su voluntad es para él [Kant], en cuanto se comporta como puro sujeto cognoscente, nada más que un hecho percibido empíricamente. [Sin embargo] [e]l orden de sucesión de la mutación podría invertirse en el [caso del barco], si el espectador tuviese fuerza para hacer que el barco marchase hacia atrás, como [en el ejemplo de la casa] lo ha tenido para mover su ojo en una dirección contrapuesta a la primera. Pues, de que la sucesión de las percepciones de las partes de la casa depende de su albedrío, quiere deducir Kant que nos es subjetiva... Pero el movimiento de su ojo en la dirección del tejado al sótano es una sucesión objetiva ... y el opuesto, desde el tejado al sótano, lo es también, tanto como puede serlo la marcha del barco (1989, pp. 135-136).

En realidad, no existe diferencia entre los dos casos; en ambos ejemplos, se trata de una sucesión objetivamente determinada. De acuerdo con Schopenhauer, Kant no habría, en la exposición de su argumento, creído que hubiese diferencia entre uno y otro caso, si hubiese tomado en cuenta que "su cuerpo es un objeto entre objetos” y que la sucesión de sus intuiciones empíricas depende de la sucesión de las impresiones producidas [causadas] por otros objetos sobre su cuerpo" (p. 136). Por consiguiente, la sucesión de intuiciones empíricas que percibe el observador son, en cualquier caso, objetivas y son independientes de la voluntad del observador.

Pero la refutación de la prueba dada por Kant no termina aquí para Schopenhauer. De acuerdo con el primero, sólo por la ley de la causalidad se puede conocer la objetividad de una sucesión. Pero el resultado de tal afirmación sería que el único tipo de sucesión objetiva que sólo se nos daría a conocer serían las causales, observa agudamente Schopenhauer. " ... el resultado de su afirmación [la de Kant] sería que nosotros no percibimos ninguna serie en el tiempo como objetiva, con excepción de la de causa y efecto ... " (Schopenhauer, 1989, p. 137). Por otra parte, como implicación directa de lo anterior, cualquier otra clase de sucesión objetiva de fenómenos sería necesariamente causal (p. 139). Sin embargo, ordinariamente podemos distinguir entre sucesiones que no son causales y aquellas que lo son. "La sucesión en el tiempo de acontecimientos que no están en relación causal es precisamente lo que se llama casualidad (*Zufall*)" (p. 139). De esta manera, v.gr., del hecho que, al salir de los predios de la Universidad de Panamá, caiga un aguacero, no se sigue de que haya una relación causal entre ambos eventos; no obstante, la sucesión entre el evento de salir de la Universidad y el aguacero es algo que puedo percibir objetivamente, sin que mi voluntad intervenga. "Igualmente, la sucesión de sonidos en una pieza de música es una sucesión determinada objetiva, y no subjetivamente, por mí, el oyente; pero ¿a quién dirá que los sonidos de la música se suceden según la ley de la causa y efecto?"(p. 138). Para Schopenhauer es inadmisible lo que plantea Kant en torno al conocimiento y la determinación objetiva de las sucesiones mediante la ley de causalidad; todos los anteriores puntos intentan mostrar las inconsecuencias de la demostración kantiana[4].

A falta de una conclusión, unas conjeturas

No pretendo proponer aquí una conclusión en firme sobre la argumentación de Schopenhauer contra Kant. Tampoco pretendo ofrecer una nueva interpretación acerca de cómo deberíamos entender a Kant para asumir su defensa de la cosa en sí o la aprioridad del principio causal, o – en términos más generales – la consistencia lógica de su Idealismo Trascendental frente a la crítica berkeliana de Schopenhauer o cualquiera. Quisiera más bien tratar de concebir el posible origen de las faltas que se le endilgan a Kant, una vez asumidas las argumentaciones de Schopenhauer. Así pues, conjeturo que las faltas de Kant podrían radicar en la multivocidad que –debido a su extrema generalidad y vaguedad– generan términos como ‘cosa’ u ‘objeto’, multivocidad que inadvertidamente recae en los términos técnicos del vocabulario kantiano, a saber, la ‘cosa en sí’ u ‘objeto transcendental’[5].

En sus particulares críticas a Kant, Schopenhauer se afana por distinguir lo fenoménico de lo conceptual. Así, Schopenhauer distingue enfáticamente entre percepción y abstracción, dos niveles en la escala del conocimiento que no deben ser confundidos. Por consiguiente, cuando hablamos de ‘objetos’ u ‘cosas’ debemos tener en claro si tales objetos o cosas se encuentran son de tipo perceptivo o abstractivo, es decir, si son representaciones (percepciones) o abstracciones (ideas, conceptos). Expresiones como 'cosa en sí', 'objeto', 'objeto en sí' u 'objeto trascendental' no parecen aclarar por sí mismas – valga la redundancia – en qué nivel o en qué clases de objetos se encuentran. Como antes ha sugerido Magee (1983) siguiendo a Schopenhauer, Kant no parece haber hecho explícito este problema. Si esto es así, podríamos afirmar que Kant ciertamente no muestra haberse planteado la multivocidad de las expresiones referidas. Ahora bien, este problema semántico trae consigo otro, de orden más general y fundamental, ¿es la cosa en sí, objeto, objeto en sí u objeto trascendental algo que existe independientemente del sujeto, o es algo que depende enteramente del sujeto?[6] Por supuesto, la respuesta que se le dé a este otro problema define las posiciones metafísicas fundamentales, el realismo e idealismo, de las cuales Kant se mantiene equidistante, según su interpretación convencional.

Pienso que lo anterior puede aplicarse perfectamente al problema de la causalidad. Si hay multivocidad en la noción de cosa u objeto en sí (y que, como hemos visto, trae consigo una postura metafísica no definida, ambivalente), lo mismo sucede en la noción de causalidad. A juicio de Schopenhauer, la causalidad es concebida por Kant como habitando entre dos mundos, el de los conceptos y el de las percepciones, sin pertenecer propiamente a ninguno de los dos. La causalidad de Kant no solo enlaza o conecta percepciones y conceptos sino también 'cosas en sí', lo cual no puede ser, si admitimos la argumentación de Schopenhauer. Es decir, la postura de Kant con relación a la causalidad oscila entre lo subjetivo y lo objetivo, entre el idealismo y el realismo. Podemos ver esto como una consecuencia de lo anterior, toda vez que se intenta conectar o enlazar causalmente el mundo del 'objeto en sí' (¿abstracción o percepción?) con el ámbito de las representaciones, los objetos concretos situados en tiempo y espacio.

Un último comentario (y conjetura también) que quisiera añadir es el siguiente. Éste tiene que ver con la observación schopenhaueriana de que Kant, en su análisis de la causalidad, no toma en cuenta el cuerpo del observador, es decir, la corporeidad o materialidad del ser humano. Kant, señala Schopenhauer, entiende al sujeto, como un 'puro sujeto de conocimiento', es decir, como un sujeto sin cuerpo. Kant no advierte la realidad del cuerpo del sujeto humano, que – en cuanto tal – es un cuerpo más entre otros cuerpos, estando, por lo tanto, determinado por las propias leyes que rigen el comportamiento de los objetos físicos. ¿Es posible que el soslayo kantiano del cuerpo, de la realidad física del ser humano, se deba también al hecho de la multivocidad referida a la noción de 'objeto', del cual 'cuerpo' es una clase específica? ¿O Kant pasó por alto la realidad del cuerpo en su análisis de la causalidad, porque le preocupaba más la libertad del ser humano que su necesidad, es decir, la causalidad propiamente hablando, característica inevitable de todo fenómeno espaciotemporal? ¿O acaso –como habrían mostrado satíricamente los relatos del Barón de Münchhausen–, los filósofos han tendido a pensarse a sí mismos como si fueran puras mentes que ignoran sus cuerpos, o cuyos cuerpos son más bien un estorbo para pensar, según aduce cierta interpretación dualista y platónica (o cartesiana) del ser humano? Para evitar que todas estas conjeturas psicológicas se conviertan en especulaciones ociosas, finalizo aquí este texto.

Notas

[1] La carta fue comentada y publicada en español en 2015 por quien fue catedrática de la Universidad de Sevilla, la Dra. Pilar López de Santa María, en el anexo de un artículo de su autoría, referido al final de este texto.

[2] Copleston (1993) nos aclara también esta problemática en Kant al señalar que ”... [Kant] empezó desde el punto de vista del sentido común de que las cosas producen un efecto en el sujeto que da origen a la sensación, la sensación siendo definida como 'el efecto de un objeto sobre la facultad de representación, en cuanto somos afectados por el objeto'. Pero este punto de vista del sentido común parece implicar la suposición de que hay cosas en sí mismas. Ya que parece implicar una inferencia de la sensación, como un efecto, a la cosa en sí, como causa ... Pero por hablar así Kant obviamente se hace vulnerable al cargo de estar aplicando la ley de la causalidad más allá de los límites que él mismo había establecido. Ha sido desde entonces una objeción común contra la doctrina de los noúmenos, considerados como cosas en sí mismas, que su existencia sea afirmada como el resultado de una inferencia causal, porque de acuerdo a los propios principios de Kant, la categoría de causa es sólo aplicable a los fenómenos"(p.270).

[3] En palabras de Magee: "Fue claro desde el principio, la mayor parte del tiempo al menos, que Kant estaba pensando el fenómeno como el imperceptible y en última instancia sustratum real del objeto, en el cual todas sus características perceptibles recaen: lo que uno llamaría el objeto objetivo, el objeto como es en sí mismo, inexperimentado por el sujeto - y, además, consideró este noumenon como la causa de nuestras sensaciones. Pero si las categorías de espacio y causalidad son características de la experiencia (y la experiencia posible) solamente, entonces, no hay ningún sentido en el cual las cosas tal como son en sí mismas pueden ser objetos de alguna clase de 'allá afuera' en el mundo, ni tampoco hay algún sentido en el cual ellos puedan originar nuestra experiencia de ellos mismos -ya que La localización en el espacio, y la causalidad, son ambas de manera similar de origen subjetivo ... Kant estaba mediatamente consciente de esta dificultad. Pero su dilema era éste: si él admitía explícitamente que los objetos físicos son las causas de nuestras sensaciones, entonces concedemos que los objetos tal como los experimentamos son cosas en sí mismas y estamos de vuelta con el empirismo precrítico de Locke…; pero si, por otro lado, dispensase totalmente la noción de un substratum objetivo en relación a la experiencia, entonces estaríamos de vuelta con Berkeley a una realidad que consiste sólo en experiencias, con la excepción de que no podríamos defender nuestra posición junto con la explicación de Berkeley o, hasta donde se puede

ver, de cualquiera- del hecho que los individuos humanos viven en un mundo compartido. Y debido a que esta última opción [la de Berkeley] fue la que más temía que fuese aceptada, Kant persistentemente se inclina hacia el primer error, aunque sin hacerse bastante explícito a él mismo lo que estaba haciendo. Los ecos de esta evasión inconsciente reverberan a través de toda su obra. En el corazón de su epistemología y ontología yace un problema cuya completa estrategia le muestra haber sido represivamente consciente de él y que, sin embargo, nunca reconoció. Estaba tratando de encontrar una solución sin formular el problema. Es decir, fue reacio a confrontar el problema hasta que estuviese seguro de que lo podía resolver- y esa fue una situación que no pudo nunca alcanzar. La doble autocontradicción de Kant en permitir las cosas tal como son en sí mismas, que aparecen en su filosofía como entidades independientemente localizadas que causan nuestras sensaciones, fue una vez aparente a todos los serios estudiosos de su trabajo, tal como lo fue la naturaleza del supuesto dilema que lo había dirigido a él. […] no sólo Schopenhauer reprende a Kant directamente por su suposición ilegítima del noúmeno como una suerte de objeto invisible, espacialmente localizado, causando las experiencias, sino que también lo reprende por su reiteración de que 'el contenido empírico de la percepción nos es *dado*' … [Igualmente] reprende a Kant […] por...su inhabilidad para explicar cómo es que el mundo percibido es el mismo para cada uno sin extender la aplicación de la causalidad a las cosas tal como son en sí mismas, o sin caer en el berkelianismo" (1983, pp. 94-95).

[4] Reiterando lo anterior, Schopenhauer señala lo siguiente. "Si la objetividad de la sucesión fuera conocida meramente por la causalidad, sólo se la podría pensar como tal, y [toda sucesión] no sería más que causalidad… Por consiguiente, si Kant tuviese razón, no se podría decir: 'Este estado le sigue a aquel', sino que 'seguir' y 'ser efecto' serían una y la misma cosa, y la proposición sería tautológica" (1989, p. 142). Schopenhauer también señala que, si la sucesión sólo nos es conocida por el nexo de necesidad causal, entonces toda sucesión real nos es dada a conocer por su carácter necesario. Sin embargo, esto presupondría que un entendimiento pudiese abarcar toda la cadena de causas-efectos, pues sólo así se constataría la necesidad de los fenómenos percibidos como sucesivos.

[5] Esta conjetura es respaldada en cierto modo por las observaciones de López de Santa María (2015) a propósito de la expresión 'objeto trascendental'. Así, señala la especialista, "[e]s menester recordar aquí que el uso kantiano de la expresión *objeto trascendental* no es unívoco, sino que presenta dos acepciones principales. En ocasiones Kant la utiliza como sinónimo de *cosa en sí*, para referirse a la causa o al fundamento desconocido de los fenómenos externos. [...]Ese es el uso que aparece, por ejemplo, en varios pasajes de los paralogismos… En otros pasajes la expresión designa el concepto de objeto en general que, a falta de la cosa en sí, es pensado por la conciencia como punto de referencia de nuestras representaciones fenoménicas a fin de garantizar la validez objetiva de las mismas. [...] Esta dualidad de significados del objeto trascendental ha sido interpretada como un descuido kantiano por algunos críticos [...] En cambio, otros autores, [...] entienden que la ambigüedad tiene un fundamento dentro del sistema del idealismo trascendental. Sea como fuere, lo que parece cierto es que esa dualidad de significados tiene mucho que ver en la disputa sobre el idealismo kantiano, pues a lo que ella apunta en último término es a la cuestión de si para Kant la objetividad del conocimiento ha de fundarse en una cosa en sí trascendente a la conciencia o en un concepto inmanente a ella (p. 19).

[6] Aprovechando una vez más las palabras de la Dra. López de Santa María citadas en el pie de página anterior, la multivocidad de las expresiones usadas por Kant (cosa en sí y objeto transcendental) "... apunta en último término … a la cuestión de si para Kant la objetividad del conocimiento ha de fundarse en una cosa en sí trascendente a la conciencia o en un concepto inmanente a ella", es decir, a las dos alternativas que he indicado en forma de preguntas.

Referencias bibliográficas

Copleston, F. (1993). *A History of Philosophy,* Tomo VI [Una historia de la filosofía, tomo VI]. Image.

López de Santa María, P. (2015). Schopenhauer y el idealismo kantiano. *Enrahonar. Quaderns de Filosofía* 55, 11-29.

Magee, B. (1983). *The Philosophy of Schopenhauer* [La filosofía de Schopenhauer]. Oxford University Press.

Schopenhauer, A. (1989). *De la Cuádruple Raíz del Principio de Razón Suficiente* (Trad. Leopoldo Eulogio Palacios). Gredos.

Schopenhauer, A. (1958). *The World as Will and Representation*. Vol. I. [El mundo como voluntad y representación]. (Trad. E.F.J. Payne). Dover.