THEORIA

Vol. 40/1 • January 2025

Established in 1952 by Miguel Sánchez-Mazas
Second Series

An International Journal for Theory, History and Foundations of Science

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

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

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

THEORIA

REVISTA DE TEORÍA, HISTORIA Y FUNDAMENTOS DE LA CIENCIA

AN INTERNATIONAL JOURNAL FOR THEORY, HISTORY AND FOUNDATIONS OF SCIENCE

ESTABLISHED in 1952 by MIGUEL SÁNCHEZ-MAZAS Second Series

EDITORIAL BOARD

Editor-in-chief: Andoni IBARRA (University of the Basque Country, UPV/EHU)
Editors: Marc ARTIGA (Universitat de València), Javier GONZÁLEZ DE PRADO (UNED)

Logic and philosophy of logic and mathematics: Eduardo A. BARRIO (Universidad de Buenos Aires)
Philosophy of language: Delia BELLERI (University of Lisbon)
Philosophy of mind and cognition: Marta JORBA (Universitat Pompeu Fabra)
Epistemology: Michele PALMIRA (Universidad Complutense de Madrid)
Philosophy of science: José DÍEZ (Universitat de Barcelona)
History of science & Science and Technology Studies: Miguel GARCÍA-SANCHO (University of Edinburgh)

Consulting Editor: Cristina CORREDOR (Universidad de Valladolid)
Managing Editor: Mario SANTOS-SOUSA (University College London)

ADVISORY BOARD

Juan José ACERO (Universidad de Granada), Ignazio ANGELELLI (†) (University of Texas at Austin), Manuel ATIENZA (Universitat d'Alacant), Salvador BARBERÀ (Universitat Autònoma de Barcelona), Nancy CARTWRIGHT (Durham University), Newton C.A. da COSTA (†) (Universidade de São Paulo), Joseph DAUBEN (City University of New York), Elías DÍAZ (Universidad Autónoma de Madrid), Javier ECHEVERRIA (Ikerbasque & UPV/EHU), Josep Maria FONT (Universitat de Barcelona), Manuel GARCÍA-CARPINTERO (Universitat de Barcelona), Philip KITCHER (Columbia University), Bruno LATOUR (†) (Sciences Po, Paris), Larry LAUDAN (†) (UNAM), Javier DE LORENZO (Universidad de Valladolid), Thomas MORMANN (UPV/EHU), C. Ulises MOULINES (Ludwig-Maxilians-Universität München), Carlos MOYA (Universitat de València), Javier MUGUERZA (†) (UNED), Roberto TORRETTI (†) (Universidad de Puerto Rico), Enric TRILLAS (Universidad Politécnica de Madrid), Nicanor URSUA (UPV/EHU), Bas C. VAN FRAASSEN (San Francisco State University)

EDITORIAL OFFICE

Editorial manager: Ainhoa FERNÁNDEZ (CALIJ, Donostia/San Sebastián, Spain)

Mailing address: CALIJ-THEORIA, Elhuyar plaza 2 (20018 Donostia/San Sebastián, Spain)
Tel.: (+34) 943 017 447. Fax: (+34) 943 015 990. E-mail: editor.theoria@ehu.eus

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

THEORIA's website

http://www.ehu.eus/theoria

Contents

Volume 40/1, January 2025, pp. 1-104

ISSN 0495-4548
e-ISSN 2171-679X

The Lullius Lectures

Articles

The Lullius Lectures

Theoria, 2025, 40(1), 5-16
https://doi.org/10.1387/theoria.27510

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Disorder

(Desorden)

John Dupré*
University of Exeter
https://orcid.org/0000-0002-7451-2127

Keywords

Classification
Disunity of science
Metaphysical disorder
Values
Sex
Gender
Race

ABSTRACT: This paper begins with some brief intellectual autobiography, recalling my first engagement with philosophy of biology. The substantive part of the paper then focuses on the plurality of possible classifications central to the theses of scientific disunity and metaphysical disorder developed in my early career. After discussing this in terms of biological classification, and introducing the reasons for thinking of classifications as typically value-laden, I discuss two sets of human classifications bearing on normatively vital questions, those around sex and gender and those involved in the distinctions between human races.

Palabras clave

Clasificación
Desunión de la ciencia
Desorden metafísico
Valores
Sexo
Género
Raza

RESUMEN: Este artículo comienza con una breve autobiografía intelectual, en la que se rememora mi primer contacto con la filosofía de la biología. La parte sustantiva del artículo se centra en la pluralidad de posibles clasificaciones central en las tesis de desunión científica y desorden metafísico desarrolladas en los comienzos de mi carrera. Después de discutir estas tesis en términos de clasificaciones científicas, y de introducir las razones para pensar en las clasificaciones como típicamente cargadas de valor, discuto dos tipos de clasificaciones humanas relacionadas con cuestiones de gran importancia normativa, las concernientes al sexo y el género y las involucradas en distinciones entre razas humanas.

1. Introduction

The organisers of the lectures on which these papers (this and Dupré 2025) are based asked that I present some reflections on my philosophical career. I shall take this literally and begin with a little intellectual autobiography.

Like so much in my life —and I think life generally— my becoming a philosopher of biology was a highly contingent matter. After interrupting my undergraduate studies with a brief and perhaps misguided idea that I would become a mu-

* ***Correspondence to:*** John Dupré. Egenis, Centre for the Study of Life Sciences; Department of Sociology, Philosophy and Anthropology; University of Exeter; Byrne House; St. German's Road; Exeter EX4 4PJ – j.a.dupre@exeter.ac.uk – https://orcid.org/0000-0002-7451-2127

How to cite: Dupré, John (2025). «Disorder»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 5-16. (https://doi.org/10.1387/theoria.27510).

Received: 08 July, 2024; Final version: 27 January, 2025.
ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

sician, I returned to Oxford to complete my studies with a firm notion that I wanted to pursue a career in philosophy. This meant that I needed to apply to a PhD programme. The problem was that in the UK one is required to say, in such an application, what the subject of one's PhD research will be; and I had no particular idea what I wanted to work on.

I recall discussing this with one of my tutors, the late Gordon Baker, and he suggested a number of topics that he thought were interesting and topical. One of these was the philosophy of biology. I knew no biology at all. Indeed my schooling included just one term of biology provided, ironically as it later struck me, by a classical German morphologist. My only recollections of this experience were the class uncomprehendingly repeating morphological liturgies, such as stem, leaf, flower or sepal, petal, stamen, carpel and, rather unpleasantly, a practical session which began with the removal of the skin from a frog's leg, an operation referred to as "debagging a frog". On the other hand, I did have a longstanding interest in classifying organisms, in my early youth breeds of cows and species of beetles, and at that later time wildflowers. As this resonated more than any other suggested topic, I made the fateful decision to take up this branch of philosophy.

Further research followed, in which it appeared that there were only three prominent living exponents of the field: Michael Ruse, and David Hull, ambitious young philosophers working in America, and (much older) Joseph Henry Woodger, a British biologist and philosopher, then often reviled as a card-carrying positivist.[1] This at least made it easy to catch up with the field. Reading Ruse and Hull (and Woodger's attempt to axiomatize evolutionary biology, surely the low point of an exceptionally interesting and productive career) it became plain that the big question in the field was whether biology was reducible to physics, perhaps via biochemistry and one or two other intermediate stops.

So I devoted myself to this question, one that now I must confess seems extraordinarily pointless. Biophysics is an interesting but specialised area of biology, but apart from that, and some reflections on the relation between the mass of an animal and the likely effects of jumping off a tall building, physics and biology have only a modest amount to say to one another.[2] The question that really drove this debate was strictly one of metaphysics rather than philosophy of science. Was there some sense in which the smallest elements of the universe were all there were? If so, as many philosophers still seem to believe, ultimate causality could only reside in these fundamental entities, and the subjects of a so-called "special science" such as biology, must ultimately be fictitious and causally inert. Perhaps this is an important metaphysical question, but it is one that is almost wholly independent of any details of biological science.

Nonetheless, I spent a great many hours reflecting on whether biology might indeed be reduced to physics. Perhaps genetics provided a link between chemistry and biology that might lead to the desired reduction? Or perhaps advances in chemistry and physics would eventually enable us to subsume biology under their growing domain? Although the answers to these questions were, at least with hindsight, quite evidently negative, reflection on why this was so led me to learn some biology, which no doubt was highly valuable in my later philosophical ventures.

2. Classification, natural kinds, and promiscuous realism

Somewhat to my surprise, my most productive pathway into philosophy of biology came from my existing, not especially intellectual, interest in classification, especially of plants. At that time, probably the liveliest topic of discussion in philosophy generally was the theory of direct reference promoted by Saul Kripke and Hilary Putnam. Putnam's (1970, 1975) application of this in the philosophy of science seemed finally to provide a solution to the vexing problem of meaning change with theory change. Then standard Fregean theories of language related meaning to descriptions of the entities referred to. But then, if theories change, so did the meanings of the words referring to their subject matters.

---

1 There were, of course, other philosophers interested in biology. But long before the internet, this was as much as I was immediately able to discover of the field.

2 No doubt physics is relevant to chemistry and chemistry to biology. But relevance is not a transitive relation and in this case the inference that physics is relevant to biology seems questionable.

Given the vast changes that had occurred in theories of, say, electrons, since they were first postulated, it seemed that the story of these changes lacked any common subject matter. Putnam proposed to solve this problem by suggesting a direct interaction between the word and the referent when a name was introduced.

So when George Johnstone Stoney proposed the name "electron" in 1891 (O'Hara 1975), for a particle that had been extensively studied by scientists for some decades, there was a fair consensus about where the entities cold be found, what could be done with them, and so on. A link was established between a word and a kind of entity. As subsequent research produced increasingly divergent accounts of the nature of that entity, this link maintained the coherence of this developing narrative. Of course, the story only works if there is a kind of entity successfully "dubbed". And the dubbing is only successful in so far as other scientists accept that the entity exists and agree to use that name.

One thing that this notorious episode in philosophy brought into focus for me was the problem of a philosophy of science so solidly focused on physics. A whole host of problems around realism, to which the question of theory change is central, are really problems for physics. The theory of evolution transformed biology, but no one worried much about whether creationists and evolutionists were talking about the same animals. A monkey is a monkey, whether created or evolved. And no one apart from globally radical sceptics supposes that monkeys might not exist —though perhaps they may not for much longer.

Nonetheless, Putnam was happy to apply the theory immediately to biology. What does the word "lemon" mean? It is all and only members of the biological kind exemplified by the lemon first dubbed "lemon" (Putnam 1970). The assumption was that biological kinds, like atoms or subatomic particles, fall into discrete natural kinds exactly one of which every biological organism belonged to. Putnam assumed that these were demarcated by an essential property, the genetic structure of the lemon. But here my experience of biological classification sounded alarm bells. Plants, particularly, are a diverse lot. Members of a species differ greatly and are often almost impossible to distinguish from close relatives. One reason for this is that they hybridise a lot. There is no more reason to suppose that there is an unequivocal genetic basis for sorting them into discrete kinds than that there is a morphological basis for this.

Putnam's chosen example of a lemon illustrates this well. Greengrocers may be very good at distinguishing the various commercially grown citrus varieties —31, according to one specialist website— and since many of these are clonal, they may even be genetically distinct. But the biological origin of these fruits remains complex and even obscure. Simplifying slightly, experts currently trace a web of relations between four ancestral species, (the papeda (*Citrus micrantha*), the citron (*Citrus medica*), the mandarin orange (*Citrus reticulata*), and the pomelo, (*Citrus maxima*), and the various current varieties of lemon (Curk *et al.* 2016).

I had the great pleasure, when living in California, of owning a Meyer Lemon tree, a truly wonderful plant, which produces delicious lemons all the year round. Or maybe not lemons? The Meyer lemon is a cross between a citron and a hybrid between a mandarin and a pummelo.

In short, and as I argued in my first published paper (Dupré 1981) if there are natural kinds in biology, they certainly are not the kinds that Putnam needs for his account of the meaning of kind terms. Putnam wanted his dubbing story to provide an account of terms in ordinary language, and as the lemon example and many others I marshalled in that paper show, this it cannot do. In ordinary language a Meyer lemon is a lemon; but not all lemons share the same genetic essence. In fact, the situation is more serious than this. If God had created a world with distinct species they might have had distinct essences until citrus breeders and others messed things up a bit. But if, as I believe, evolution did the job, the project of assigning essences even to the biological kinds of science just won't work.

Take a tortoise and an oak tree. Trace the evolution of the tortoise back to its last common ancestor with an oak tree, something all evolutionists believe to be possible. Now go forward along the lineage that leads to the oak. We now have a series of species leading from lemon to oak in a fairly smooth series of changes. There are no (or anyhow few) sudden jumps at which a species with one essence is replaced by a new species with another. It is clear that an evolved biological world does not contain essences that separate every kind of organism from every other. And indeed, we now know that

transfer of genes between species is not just something done by God-playing plant breeders, but happens all the time in nature. The biological world is much messier than Putnam assumed.

One thing that this excursion into taxonomy led me to see, then, was that popular theories of natural kinds with Lockean real essences dividing the things of the world into perfect classifications, made little sense in biology. But this opened the door to a surely more important insight: classification is always relative to a purpose. In my earlier paper I made this point mainly in relation to the differences between ordinary language and scientific language. Chefs, carpenters, gardeners and many others have reasons to distinguish kinds of organisms, and their interest need not coincide with those of scientists. Thus, for instance, for a carpenter cedar is a kind of wood with distinctive properties, notably its moth-repellence and rot-resistance, the former property making it popular for lining chests and closets. In Europe, this wood is generally from the genus *Cedrus*, a kind of pine. In the US, the same function is often served by the Eastern red cedar, *Juniperus virginiana,* a species of Juniper. The Western red cedar, *Thuja plicata*, a species of cypress, is, like its other namesakes, strongly rot resistant, and particularly suitable for many outdoor uses (Dupré 1981).

A number of philosophers have objected to this line of argument, that natural kinds were always intended as the kinds of science, and little, apart from Putnam's account of ordinary language terms, follows from the failure of the mundane aims of ordinary folk to sort the world in alignment with the kinds of science. Science has ultimately just one goal, discovering the truth. As it happens, however, and as has been increasingly apparent in investigations by philosophers of biology over the last forty years, the more technical scientific concept of a species presents similar difficulties. Current estimates suggest there are over two dozen species concept circulating among biologists, some based purely on evolutionary history, some on ecology, some on physiology, some on genetics. There is no reason to suppose that these will always coincide and, while some will no doubt prove more useful than others, no reason to think that one will serve all legitimate biological purposes. The solution to the so-called species problem, the problem of how to define the species, is that there is no solution to the species problem. Many different definitions serve different purposes (ecology and phylogeny, for instance) and are suitable for different domains of life (see, e.g., Dupré 1993, ch. 2).

I named the general picture to which these observations have led me "promiscuous realism". There are many ways of dividing the entities in the world into kinds (promiscuity), and these kinds may often overlap one another. But there is no reason to deny that these kinds —even, I argue, the kinds of ordinary language— are perfectly real. They reflect properties, often clusters of properties, that truly pertain to the objects to which they are applied, and identifying something as belonging to such a kind is often valuable for predicting how it will behave and the purposes, practical or intellectual, that it may serve.

Once essentialism is firmly dismissed, such a position is immediately appealing. If there is no privileged set of properties, the essential properties, that determine how things should be classified, why assume that any mode of classification will serve all the purposes for which we classify. Think, again, of biological species. The most ancient reason for classifying is simply to enable the recording and storing of information. If I see one biological entity devouring another, it is helpful for all sorts of reasons to be able to say that I saw, for instance, an orca eating a seal. There is a fair chance that this is a characteristic activity of things of that kind, the orcas; I should keep my pet seal away from orcas. This motive does point to the value of a general taxonomic system that aims to assign every organism to exactly one species. But the motivation for such a system is entirely pragmatic; it cannot hope to reflect some real structure of the living world.

3. Against reductionism

In this early work, reflection on the failure of essentialism led me also in a different direction, back to my concern with reductionism. Although the consensus was beginning to crack in the 1970s (Fodor 1974), even in the 1980s physicalist reductionism was still a widely assumed position. This was the idea that all science was really physics, that ultimately the findings of others would either be found in adequate or be interpreted in the language of physics. How was this in-

terpretation imagined? The concepts of the "higher" level science, that is, the science that dealt with more complex objects, must be identified with arrangements of the objects of fundamental physics and the behaviour of the more complex objects deduced from the laws that apply to these physical parts. This seemed plausible enough for the reduction of chemistry to physics —atoms were understood as complexes of more fundamental particles, and molecules as arrangements of atoms— and the reduction of biology to chemistry seemed at least to have been begun by work in genetics and other parts of molecular biology.

But note that the proper arrangement of components looks exactly like the essential property of the higher level entity. If the properties of water, conceived as a mass of H_2O molecules, is to be inferred from the physical properties of its parts, then every H_2O molecule must have precisely the right structure, its essential property. So threats to essentialism looked likely to provide problems also for reductionism.

Perhaps at higher levels, such as the biology of organisms, eliminativism is the way to go. These sciences cannot be strictly true because their kinds are too heterogeneous to provide proper subjects for scientific laws. If organismic biology is irreducible so much the worse for organismic biology. But essentialism was beginning to look problematic even at much lower levels. Consider again water. The pure water made up only of identical H_2O molecules was nowhere to be found in the real world. Even allowing that we are not interested in sea water, pond water or even tap water, does a proportion of heavy water, deuterium oxide, disqualify a sample from being strictly water. But even samples of pure H_2O are problematic. It is now clear that hydrogen bonds between water molecules can form much larger transient structures within liquid water, which are important in explaining the highly anomalous properties of the fluid, and that also have important implications for biochemistry. These transient structures of water are crucial, for instance, in explaining protein dynamics and protein-protein interactions (Raschke 2006).

Considerations of this kind have convinced most scientists and many philosophers that reductionism as a practical exercise is wholly infeasible. Detail emerges in complex systems that could never have been foreseen from the perspective of so-called fundamental physics. But philosophers, at least, have generally held on to the idea that "in principle", something often spelled out in terms of what would be possible for the infinite mind of God, everything might be a deductive consequence of the laws of physics and the arrangements of physical parts. This may be understood most simply merely as the idea that fundamental physics is complete. The movements and behaviour of every fundamental physical particle is fully determined, it is said, by the arrangement of the physical particles and the laws that govern their behaviour. It may be practically impossible to explain or predict the behaviour of an animal from the laws of physics. But if the animal is composed entirely of physical particles, and the position of all of these is determined by physical laws, it seems that at least the position of the animal and any changes in that position must be a necessary consequence of the laws of physics. This claim has come to be expressed by the idea that higher level changes "supervene" on the underlying physics. Biological facts or mental facts (the main home of this thesis) depend wholly on underlying physical facts (Kim 1984).

Supervenience claims remain, pretty much, orthodoxy among philosophers. For my present purposes, it is perhaps sufficient to note that supervenience represents precisely the recognition that actual reduction is impossible. But I do think that, separated from its empirical origins in reductionism, supervenience quickly becomes hard to defend. I'll mention just one reason. A first intuition about supervenience is that the properties of, for instance, an animal, supervene on the physical structure of the animal. But then we must observe that the behaviour of the animal will often depend on features of its environment, such as the presence of food, possible mates, or opportunities for play. So the supervenient base, the physical material on which the animal's behaviour depends, must include much beyond the animal itself. And these elements of the environment will themselves often have a supervenient base that stretches beyond themselves. This suspicion has been reinforced in recent years by the thesis of externalism in philosophy of mind, that the mind extends beyond the confines of the minded individual. My memory, for instance, is now based as much in my computer and my phone as in my head. It begins to look as if there might be no limit to how much of the world may need to be included to exhaust the parts whose physical properties may be relevant to our target of interest. And this undermines any hope that there could be anything approximating empirical verification of supervenience.

This leaves the defender of supervenience with nothing but the belief in the completeness of physics. And given the impossibility of finding empirical evidence for this belief, it looks increasingly like an unmotivated dogma. In reality, physical laws are discovered and tested in very specific and highly controlled laboratory circumstances. As Nancy Cartwright (1983) has famously argued, it is quite unclear how these laws apply, if at all, to the messy conditions of the real world. Why should we suppose that the laws that apply to an electron in a cloud chamber apply just as well to an electron embedded in the complex metabolic processes of an organism?

The best that can be said for the completeness of physics is that it is a purely speculative thesis, that belongs to abstract metaphysics more than to science. As a believer in the kind of metaphysics —naturalistic metaphysics— that is answerable to scientific findings about the world, I am highly sceptical of such a thesis. But at this point it may be more useful to move on to something much closer to the science, with a return to biology. One thing that historians and philosophers of biology have come to a considerable degree of consensus about is that biologists generally have little interest in the formulation of laws[3]; their practice revolves rather about the construction of models. These models are often quite unique to a very specific phenomenon or problem.[4]

The realisation that biology deals in models, or families of models, rather than laws is a crucial one for a simple reason. Whereas laws are considered to be true, even universally true, models are generally seen not even to aim at exact truth. They are idealisations and/or abstractions, perhaps even fictions. The point is not to say exactly how the world is but to identify aspects of the complexity of the world that have a strong, perhaps even decisive, effect on what happens. But it is generally understood that other factors may, in a specific case, intervene to prevent the expected outcome. And it may even be that nothing particular intervenes but the outcome fails to occur; nature is increasingly recognised to be, to some variable extent, irreducibly probabilistic. In the actual science, then, in so far as this consists of models rather than laws, we do not even have candidates for the potential reduction to (the laws of) physics.

4. Values in science

I want now to go back to kinds, and some important ways in which the abandonment of the essentialist view of kinds really matters a great deal. I have argued that there are many ways of classifying the world into kinds and the correct, or best, way of doing so depends on the purposes for which our classification is intended. There is no single, objectively best, classification that science can provide for us. But this opens up a whole range of normative questions about science. What are the goals underlying specific classificatory schemes? And may there be social, political or ethical values implicit in the way we do science? After drawing together my arguments on essentialism and reductionism in my 1993 book, *The Disorder of Things*, these questions were my central philosophical concerns for the next decade or so.

Let me begin with the general question of values in science. This has taken many fascinating directions in recent philosophy of science. We have an important body of work on inductive risk, drawing attention to the necessity of weighing up the costs of getting our scientific conclusions right or wrong (Douglas 2000). Finding out experimentally whether a chain reaction might propagate through the oceans is a bad line of research to pursue, even if the undesirable possible outcome, the total destruction of our planet, is judged as having a very low probability. More difficult questions of this sort arise in social contexts. If, for example, we feel the need to explore the possibility that certain groups of people are less intelligent than others, we should consider carefully the costs of getting a false positive answer.

---

[3] Whether there might nonetheless *be* laws is a trickier question, since it depends what you mean by "law". John Beatty (1995) robustly rejects biological laws; Sandra Mitchell (1997) dilutes the idea of a law sufficiently to make biological laws possible. Almost everyone agrees that the universal, necessary, laws of nature once assumed to be central to all science, play no important role in biology.

[4] There is a large philosophical literature on models. For a good overview see, e.g., Downes 2021.

Another major research program in this direction has been the concern with epistemic justice, mainly derived from Miranda Fricker's (2007) influential book of that name. This concerns the unjust exclusion of many groups of people from the production and application of knowledge.

I shall return to epistemic justice in a moment. But there are more immediate applications of the plurality of possible classifications to values in science. Consider some fundamental concepts in macroeconomics, for example inflation. The casual observer might imagine that the change in price level was just an objective fact about an economy. But as economists are well aware, it is nothing of the sort. Not all prices go up or down at the same rate, so the inflation rate will be measured differently according to what so-called "basket of goods" we look at. Could we not look at all of them, weighted according to how much is spent on them? Something like this could have its uses, but it would not solve the problem. Not everyone, probably no one, spends exactly the average amount on every kind of good. And these differences are systematic in certain ways. Luxury yachts and private jets have a disproportionate effect on such a measure, disproportionate in that their high cost multiplies their effect but also makes their price irrelevant to all but a tiny minority of consumers. In general, however inflation is measured for a whole economy, it will be different for different consumers, or categories of consumers, depending on what exactly they buy[5].

If the price of such luxury items is falling, but the price of essentials such as food and housing is rising, poorer people may experience a higher inflation rate than the general average for a nation. Their wages, pension or social security benefits, when indexed to this average rate, will then leave them poorer. This is a clear case of epistemic injustice: the results of scientific measurement, in which the poor typically have little voice, turn out to disadvantage them.

I don't propose to speculate here on whether there are malign motives at work. The general point is that there is no way of measuring inflation that does not involve making decisions that affect different people differently. Values —whose goods matter most— are impossible to exclude from the measure. Inflation, in short, is not a natural kind. Similarly with a global measure such as gross domestic product, or GDP. This is generally supposed to measure the productive activity of an economy. This is hard to measure. The normal procedure is to measure such activity in terms of the money paid for the results of this activity. Since GDP is a policy target of most governments, how different activities will affect this quantity affects the ways that governments encourage and reward different activities. But suppose instead of measuring exchange value, we tried to measure GDP in terms of use value. One might imagine that lower use value might be attached to luxury yachts, designer clothes and perhaps nuclear weapons and cigarettes, than to food and housing. Even true believers in the all-knowing powers of markets should know that well-known market distortions caused by widespread monopoly power limit the benign effects of this supposed omniscience[6].

More important still is a problem with GDP eloquently explained decades ago by Marilyn Waring (1988). Many kinds of productive work happen outside the monetary economy altogether. Notable among these is domestic work such as childcare, cooking, small scale food production and care of the aged. In most countries this work, invisible to GDP, is both unpaid and overwhelmingly performed by women. The standard procedures for measuring GDP are, therefore, a massive source of epistemic justice to a vast proportion of the human population. And as Waring explained in some detail, this has real and massive effects to the disadvantage not only of women, but to children and often even to men. It would, admittedly, not be easy to measure GDP in terms of use value. But the attempt to do so, making it impossible to conceal the highly political, value-laden nature of the concept, would surely be salutary. Note finally, that I am not saying that there is anything epistemically wrong with measuring GDP in terms of exchange value. It is part of one scheme with which we can provide conceptual order on the chaotic array of economic activity. But we should recognise that it is not the only such scheme; that it carries with it particular goals and particular values; and that it has real consequences on the way we all live.

---

5 The relevance of this example to values in science is discussed further in Dupré 2007.

6 I do not mean to imply that they actually do. Recent leaders of the British Conservative party, most spectacularly and disastrously the ephemeral Prime Minister, Liz Truss, provide some obvious counter-examples.

5. Sex and gender

Another classificatory question has become massively contentious in recent political debate, the definitions of sex and gender. It is hardly within the scope of this paper to resolve these questions here, and I shall not try. But at least the beginning of wisdom is to recognise the purpose relativity of classification. If we are interested in general questions in evolutionary theory, for example, there is no question that sex, generally defined as production of large or small gametes, eggs and sperm, is an unavoidable classification. We should note first, however, that not all organisms have sexes. Second, and more important, as Paul Griffiths (2021) has argued in recent work, sex is not an essential property of any organism understood, as it should be, as an entire life cycle. At many stages of their life, organisms do not produce gametes, large or small. Some organisms, for example the Eastern Blue Groper (*Achoerodus viridis*), produce small gametes at an early stage in their life and large gametes later on. It also changes from brown to blue when it comes to satisfy the biological definition of being male. Ocellaris clownfish (*Amphiprion ocellaris),* begins life as a male. When the female in a community dies, the highest ranking male transforms into a female. Sex is, nonetheless, in these and similar cases, an important biological property that organism possess at stages of their life cycles. Griffiths (2021) argues that an immature organism that has yet to develop the ability to produce gametes is, in this sense, neither male nor female.

Humans are not classified by sex because of our overwhelming interest in evolution or gamete production. There are many institutional functions that are served by this classification, but we may very reasonably inquire which of these are important and what the appropriate criteria are for applying them. That the latter question is only contingently related to the biological distinction just discussed is clear from the fact that a substantial proportion of humans, for various reasons, do not at any particular time have any capacity to produce gametes.

Of course, most people think of sexual difference in terms of genital morphology and secondary sexual characteristics such as beards and breasts, rather than gamete size. Here it is important to be aware that sexually dimorphic development is a complex developmental process with a huge array of different outcomes, loosely clustered around what are thought of as the paradigmatically male and female (Dupré 2017). Within and beyond these clusters there is great diversity and no universal correlation between the various genetic, hormonal and physiological features that constitute the loose clusters. When it comes to the social and behavioural characteristics more commonly thought of as constituting gender, these are enormously diverse. The degree of clustering around norms of employment, domestic work, styles of dress and so on is now seen to be overwhelmingly due to cultural norms, and in a society that values individual freedom of choice, there is no possible reason to object to the wishes of the many gay, non-binary or transgendered people who prefer to diverge from these norms. I shall say a bit more about this in my second lecture.

6. Race

My final example will also be a controversial one, though a declining controversy from a biological perspective, the question of racial classifications. No biologist now thinks that the human species has well distinguished subspecies, and even the biological term "race" which has no formal definition in taxonomy, has no clear application to *Homo sapiens*. The species is genetically unusually homogeneous, and reproductively panmictic, that is to say there are no biological barriers to interbreeding between any identifiable subpopulations.

This is surprising to many people, as they are often brought up to classify most people immediately as belonging to specific racial groups, and experience this as easy to do[7]. What are we doing when we classify people, on the basis of a few

[7] There is a substantial and growing literature on the philosophy of race. For a survey and extensive further citations, see James and Burgos (2024).

superficial, but to us salient, physiological traits? There is a long tradition of racial science, almost all of which has been thoroughly discredited. But the conviction of the reality of race has been reinforced —to many people if not biologists— by the recent announcement that races can reliably be distinguished by genetic tests[8]. What are these tests testing, if not distinct races?

Why do people from equatorial climates generally have dark skin? This is quickly answered. Melanin in the skin, the chemical which causes dark coloration, protects from sun damage. Why do people from cooler climates generally have lighter skin? Again there is a readily available answer. Humans use sunlight to make vitamin D, and melanic skin is less able to use sunlight. Strong climate-based selection has rapidly optimised skin colour as human populations moved from hot to cold climates and vice versa (Jablonski and Chaplin 2000). Such adaptation to local conditions is found throughout life. Often, especially in plants, such local adaptation is purely developmental. For example, the North American lake cress, *Rorippa aquatica*, produces entirely different leaf shapes according to whether or not it is submerged in water, and also in response to increasing temperature. Animals tend to be less developmentally plastic, but genetic and epigenetic switches can allow rapid evolution of environmentally sensitive traits, and this seems to be the case for human skin colour.

The genetic tests just mentioned do not, of course, merely confirm that people have genetic tendencies to more or less melanized skin, something that can be done without much specialized equipment. Rather they claim to sort people into distinguishable populations of some kind. The first thing to note about this research is that the programme used by Rosenberg and colleagues (2002) to analyze genetic data in this way, *structure*, does not tell us how many human populations there are. Rather, it is given a number, and on the basis of a measure of genetic similarity it provides the best clustering it can find into that number of groups. The genetic data it uses does not track functional genes for skin colour or hair texture, but what are believed to be random mutations on non-functional genes. Since these are (or so it is believed) not subject to selection, they accumulate over time, and thus measure the distance in time that groups of humans have been reproductively separated from one another. It is thus a way of exploring human migration patterns.

As the number fed into the programme increases, new such groups are added by splitting existing groups. N = 2, separates Africa and America; n = 3 separates Eurasia from America; n = 4 separates East Asia from Eurasia; and n = 5 separates Oceania from East Asia. This all fits quite well with traditional views of continentally based racial groups. N = 6, however, distinguishes the Kalash people of Northwest Pakistan, population about 4,000, and not an obvious candidate for a distinct race.

As I noted, this methodology measures migratory history rather than any significant differences between the groups distinguished. As the authors acknowledge, the vast majority of variation between humans exists within not between these populations. And anyhow, the genetic features used are specifically chosen not to mark functional differences. Nonetheless, it is striking that these markers do in fact still provide a fairly reliable criterion for deciding the continental origin of an individual. Why is this?

As I have mentioned, there are no systematic biological barriers to interbreeding between the groups of people distinguished by *structure*. The human species is highly migratory, as we all hear much too much about nowadays, and major migrations of the past can be traced through these methods. But given the possibility of interbreeding, it is perhaps surprising that these marks of ancestry have not been more thoroughly distributed through the population.

Some of my ancestors a mere 350 years ago, were Huguenots, who migrated en masse from France in 1685 to avoid religious persecution. But in the intervening dozen or so generations the amount of interbreeding with, in my case, German and English ancestors makes it unlikely that I could be genetically distinguished as a Huguenot. The ancestor who bequeathed me my family name was only one of a few thousand ancestors since that time, though perhaps a few others

[8] The most famous such report, and the basis for discussion below, is Rosenberg *et al.* (2002).

left France with him. Some groups of people have been a lot less mobile and liable to dispersal than others, and distinguishing genetic markers will be proportionately more concentrated.

But probably most significant is something quite different. I said there are no biological barriers to interbreeding between human groups, but there are plenty of social barriers. Broadly speaking, these are what we call racism. I use this term here very broadly, not in a necessarily ethical sense. There are more or less problematic ways of retaining the identity of human groups by preferential inbreeding, all of which require making distinctions that are (broadly) racial. Nonetheless, without any such preferences, and given the level of human mobility that currently exists, these markers of ancestral geography would dissipate very rapidly. For this reason, I conclude that far from racism being a response to the perception of the objective fact of racial difference, racial differences are largely a consequence of racism. Race does not cause racism; racism causes race.

This leads me back to my main theme. Race turns out to be a wonderful illustration of the necessity of asking what a classificatory system is for. Historians have long explored the ways in which racialist ideas have been used to justify appalling exploitation of people from other parts of the world. Whatever the intentions of early racial classifiers, this seems to have been the main purpose that it served. And it is not hard to argue that this is the purpose that it still serves, though the issue is complicated by the argument that we now need racial classifications to identify the people to whom reparations are owed for past injustice. But even if we accept this last point (and, less probably, act on it), the ultimate goal should be the end of racial categories even if there are processes of reparation that will be needed before this goal can be properly obtained. Racial kinds are not interesting biological kinds, they are social kinds; and social kinds that have been used for almost exclusively malign ends.

7. Conclusion

It is time to summarise. The philosophical journey I have been sketching began with a question seemingly quite internal to science, even a quite technical question. What is the relation between the sciences of complex things and the sciences of the simpler parts of which they are composed? A commonly held view at the time my work began was the answer summarised as reductionism: ultimately we should explain the behaviour of anything in terms of the behaviour of its parts. This led me to the consideration of how we begin to impose epistemic order on things at any level of organisation, the theory of classification. The standard reductionist view was that there was an objective answer to this question, the scientific search for the natural kinds into which the world was organised. But a careful investigation of biological classification showed there were no such natural kinds. Kinds were only natural in so far as they best served particular goals of enquiry. Classifiers look to maximise homogeneity of their kinds not absolutely, which is impossible, but in respect of characteristics relevant to a particular kind of question. In biology, such questions might be evolutionary, ecological, physiological, genetic and so on, and there is no guarantee that these will coincide. And, for that matter, the interests of the gardener, the chef, the carpenter or the forester may also quite legitimately motivate their own distinctive modes of classification.

This observation demolished one central plank of the reductionist programme, since it showed that we did not have the natural kinds that would provide the links between sciences in the hierarchy of complexity. There is no unique answer to how we should define a human or a banyan tree in terms of its smaller components. But the realisation of the purpose relativity of enquiry points to something much larger that takes us beyond the internal understanding of science.

If classification, even scientific classification, is relative to human purposes then science itself must be shaped in fundamental ways by our interests. One influential idea that cannot survive this insight is the ideal of science as wholly value free. But if this is right, then we must also look at the ways that science influences human life, and does so in ways that depend on the values that were incorporated in the production of science. An obvious consequence of this is that

science will tend to reinforce and validate the social and ethical ideas that have been fed into it from the outset. The deeper and more significant the social ideas involved, the stronger this reinforcement is likely to be, so that the science based on concepts —classificatory concepts— that most generally shape our lives, such as the science of sex and gender or the science of race, are most seriously in need of critical scrutiny. This, I am pleased to note, is something that has become an increasingly central part of the work of philosophers of science in the last few decades.

This talk has also illustrated a broad feature of my work that remains controversial within the philosophy of science, the deep interconnections between science and metaphysics. I started with an idea that has traditionally been considered a question central to metaphysics, whether the world is composed of natural kinds defined by essential properties, and used my answer to reach some broad conclusions about science. But this is not the metaphysics traditionally conceived of as an a priori activity that precedes science, but a metaphysics grounded in reflection on the findings of science; the conclusion about natural kinds and essences was based on exploration of scientific (and indeed extra-scientific) empirical knowledge. Just as science is partly grounded in human values and human values are further shaped by science, so with metaphysics. The metaphysics I am concerned with, sometimes called naturalistic metaphysics, is grounded in science, but such a metaphysics can also redirect and improve the way we do science.

In the next paper I turn to a more ambitious metaphysical thesis that has shaped the last two decades of my work, the claim that we live in a world not composed of discrete and sometimes interacting things, but of inextricably intertwined processes. Again, I shall hope to show how this thesis best provides coherence for the findings of the various sciences, and how at the same time it reshapes much of how we should do science, and how we should do science in ways better able to serve our human purposes. To quote the concluding sentence of my 1993 book, "like other human products, the only way [science] can ultimately be evaluated is in terms of whether it contributes to the thriving of the sentient beings in this universe". I hope that the ideas summarised in this paper and the next can be seen have been shaped by the hope of making a contribution, however small, to this goal.

Acknowledgements

This and the following paper (Dupré 2025) are based on the Raimundus Lullius lectures given in Oviedo, Spain in July 2024 at the 11th annual meeting of Spanish Society for Logic and Philosophy of. Science. I am very grateful to the organisers of the conference, Cristina Corredor and Javier Suárez, first for inviting me to give the lectures, and then for making my visit to Oviedo so memorably enjoyable.

REFERENCES

Beatty, J. (1995). The evolutionary contingency thesis. In G. Wolters and J.G. Lennox (Eds.). *Concepts, theories, and rationality in the biological sciences* (pp. 45-81). University of Pittsburgh Press.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Curk, F., Ollitrault, F., Garcia-Lor, A., Luro, F., Navarro, L., Ollitrault, P. (2016). Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. *Annals of Botany, 117*(4), 565-83.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of science*, *67*(4), 559-579.

Downes, S.M. (2021). *Models and modeling in the sciences: A philosophical introduction*. Routledge.

Dupré, J. (1981). Natural kinds and biological taxa. *Philosophical Review*, *90*, 66-91.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press.

Dupré, J. (2007). Fact and value. In H. Kincaid, J. Dupré, and A. Wylie (Eds.). *Value-free science: Ideals and illusions* (pp. 27-41). Oxford University Press.

Dupré, J. (2017). Postgenomic perspectives on sex and gender. In D. L. Smith (Ed.). *How biology shapes philosophy: New foundations for naturalism* (pp. 227-246). Cambridge University Press.

Dupré, J. (2025). Process. *Theoria. An International Journal for Theory, History and Foundations of Science*, 40(1), 17-28.

Fodor, J.A. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese*, *28*(2), 97-115.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Griffiths, P.E. (2021). What are biological sexes? *PhilSci Archive*. https://philsci-archive.pitt.edu/19906/1/Griffiths%20-%20What%20 are%20biological%20sexes.pdf .

Jablonski, N.G., and Chaplin, G. (2000). The evolution of human skin coloration. *Journal of human evolution 39*(1), 57-106.

James, M. and A. Burgos (2024). Race. In E.N. Zalta & U. Nodelman (Eds.). *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition). URL = <https://plato.stanford.edu/archives/spr2024/entries/race/>.

Kim, J. (1984). Concepts of supervenience. *Philosophy and phenomenological research 45*(2), 153-176.

Mitchell, S.D. (1997). Dimensions of scientific law. *Philosophy of science 67*(2), 242-265.

O'Hara, J.G. (1975). George Johnstone Stoney, FRS and the concept of the electron. *Notes and Records of the Royal Society of London*, *29*(2), 265-276.

Putnam, H. (1970). Is semantics possible? *Metaphilosophy 1*(3), 187-201

Putnam, H. (1975). The meaning of "meaning". *Minnesota Studies in the Philosophy of Science 7*, 131-93.

Raschke, T.M. (2006). Water structure and interactions with protein surfaces. *Current opinion in structural biology*, *16*(2), 152-159.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., and Feldman, M. W. (2002). Genetic structure of human populations. *Science 298*(5602), 2381-2385.

Waring, M. (1988). *If women counted: A new feminist economics*. Harper & Row.

John Dupré is Professor of Philosophy of Science at the University of Exeter (UK), where he directed Egenis, the Centre for the Study of Life Sciences, for 20 years (2002-2022). He specializes in the philosophy of science, with a particular interest in the philosophy of biology.

Address: Egenis, Centre for the Study of Life Sciences; Department of Social and Political Sciences, Philosophy, and Anthropology; University of Exeter; Byrne House; St. German's Road; Exeter EX4 4PJ. Email: J.A.Dupre@exeter.ac.uk – ORCID: https://orcid.org/0000-0002-7451-2127

Theoria, 2025, 40(1), 17-28
https://doi.org/10.1387/theoria.27512

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Process

(Proceso)

John Dupré*
University of Exeter
https://orcid.org/0000-0002-7451-2127

According to the theory of the correlatives, the nature of a being is something defined by its activity. Therefore, being and activity are inseparable and identified.

Stanford Encyclopedia of Philosophy on Ramon Llull (Priani 2021)

Keywords

Process ontology
Organism
Metaphysical disorder
Human condition
Personal identity
Free will
Sex
Gender
Race

ABSTRACT: In this paper I discuss the process ontology that has been the central focus of my research for almost 20 years. I explain what this is, and illustrate how it applies to biology through the example of the organism. I also aim to show how naturally process ontology fits with the disordered world I described in the preceding article. Finally, I show how process philosophy illuminates a number of topics relating to the human condition, including personal identity and freedom of the will, and provides a deeper understanding of the issues around human classification, notably by sex and gender and by race.

Palabras clave

Ontología de procesos
Organismo
Desorden metafísico
Condición humana
Identidad personal
Libre albedrío
Sexo
Género
Raza

RESUMEN: En este artículo discuto la ontología de procesos que ha constituido el foco central de mi investigación durante casi 20 años. Explico en qué consiste e ilustro cómo se aplica a la biología a través del ejemplo del organismo. También me propongo mostrar cómo la ontología de procesos encaja de forma natural con el mundo desordenado descrito en el artículo que precede a este. Finalmente, muestro cómo la ontología de procesos ilumina una serie de cuestiones relacionadas con la condición humana, incluyendo la identidad personal y el libre albedrío, y proporciona una comprensión más profunda del problema de las clasificaciones humanas, en particular por sexo y género y por raza.

* ***Correspondence to:*** John Dupré. Egenis, Centre for the Study of Life Sciences; Department of Sociology, Philosophy and Anthropology; University of Exeter; Byrne House; St. German's Road; Exeter EX4 4PJ – j.a.dupre@exeter.ac.uk – https://orcid.org/0000-0002-7451-2127

How to cite: Dupré, John (2025). «Process»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 17-28. (https://doi.org/10.1387/theoria.27512).

Received: 10 July, 2024; Final version: 27 January, 2025.
ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

1. Introduction

In the last lecture (Dupré 2025b) I described some of the thinking that led me to see the world as largely disordered, a world in which the perfect order sometimes assumed by philosophers was, far from being omnipresent, a rare and fragile thing. Aspects of partial order made possible a variety of different ways of categorising, these categorisations themselves being ways of imposing order. I also explored some of the more value-related implications of this view.

My general assumption at this point was that this view of the world was simply an empirical one; it was what careful observation and our best science showed to be the case. It neither included nor required any deeper explanation of why it was like that. It just was. However, for a period now approaching 20 years I have been drawn to a more systematic understanding of the chaotic universe.[1] I believe that the reason that the disordered, chaotic universe seems so counterintuitive to many is that we are committed to a fundamentally misguided metaphysics. Whereas the world view I have been criticizing —what I sometimes call, abusively, the billiard ball view— assumes a world of discrete and autonomous things, I now see the world as composed solely of processes. While the things in the billiard ball world are autonomous, have reasonably clear boundaries, and generally possess essential properties that make them the kinds of things they are, none of this is true of processes. A process is always only partially and temporarily stable, as its stability depends on activity that sustains it. In fact, I define a process as requiring change for its continued existence.

A world of process, then, is a world of constant change. In the words attributed in antiquity to Heraclitus, 'Everything Flows'. Much of this process, like the swirls of air in the wind or the Brownian motion of particles in a fluid is directionless and chaotic. The challenge for a process philosopher is to explain the appearance of coherent and persistent patterns in this flow that has led so many philosophers to propose a world of things. The starting point for my processualism comes from the observation that such explanations are what a large part of biology is concerned with.

My main concern has always been with the life sciences, and they will be my focus today. However, I shall say just a few words about physics. Although everyone knows that the real world described by contemporary physics is nothing like the billiard ball world perhaps imagined by Laplace, the deterministic world, entirely determined by the exceptionless laws of microphysics still has many adherents among philosophers.

About microphysics, I shall say only that the idea of billiard ball-like atoms is far behind us. Nowadays the fundamental constituents of reality are seen as waves, multi-dimensional strings, and suchlike. These strike me, and some I talk to more expert in physics, as much more process-like than thing-like. But for reasons of both space and competence, I won't pursue that debate. If I am wrong, then the fundamental nature of stuff is very different from the parts of the universe I am more directly concerned with. Fortunately, however, this won't greatly alter much of what I want to say about biology. But perhaps more immediately relevant than microphysics is cosmology. Current orthodoxy is that the universe began with a bang.[2] Of course there are worries about what happened before this. Perhaps there has been an infinite cycling of similar universes. But here, fortunately, we are at a level of speculation beyond the pay grade of a mere philosopher, and I leave that to the physicists. After the big bang, at any rate, we are told of an inconceivably short period of time of almost complete chaos, and eventually, after a long stretch of around one ten thousandth of a second stable items such as protons and neutrons begin to emerge. After a few 100,000 years some of these captured electrons and

---

1 I first attempted to articulate a process view of biology in my Spinoza Lectures at the University of Amsterdam, published as Dupré 2008. More recent elaborations can be found in a number of essays in Dupré and Nicholson (2018), especially the extended introduction, and in Dupré (2025a).

2 See https://en.wikipedia.org/wiki/Timeline_of_the_early_universe. I do not rest anything on the details of the current story. For example, the theory of inflation, which claims that the early universe expanded by a factor of the order of 10^{26} over a time of the order of 10^{-33} to 10^{-32} seconds, is not easy to take seriously for the outsider to cosmology, or anyhow for me. That the history of the universe was a process during which various entities at many different spatial and temporal scales gradually emerged and stabilised seems much less controversial, and is sufficient for present purposes.

formed the atoms we know today. Within a couple of million years the gravitational attractions between these atoms had led to the formation of the first stars. And so on.

My point in referring to this story is just to note that it is a story of emergence of more or less stable entities from chaotic and formless processes. Some atoms have expected life spans exceeding the expected duration of the universe, and are, therefore as stable as it gets. Stars have life cycles, though these last for billions of years. From our point of view they are pretty stable most of the time, but of course this stability is a mark of the vast quantity of energy, ultimately exhausted, that maintains them in the state we now know them. But nothing is absolutely stable, and no thing has existed for ever. Biological evolution, though, certainly generating vastly more complex structures than any that preceded it, is very much of a piece with this story of the self-assembly of more complex forms from processes in which no such forms existed. The question of how a set of eternal things governed by precise and universal laws, might eventually have led inexorably from pure chaos to the structures we find today is probably best answered by appeal to an omnipotent creator. But fortunately for those who struggle to imagine such a creator, this is not a question we have to answer in the world of process.

2. Organisms

My view of this unifying picture of the universe did not start from cosmology, however, but from biology, so that is where I shall now turn. The natural place to start in biology is with the organism. Though it is not at all easy to define an organism, more people are confident, to borrow a phrase famously used by a US judge to describe pornography, that they know one when they see one. If you are standing by an impressive tree, or there is an elephant in the room, it seems easy enough to identify the organism. On the other hand, it is easy to show that things are not so simple. Suppose you are standing instead, by a famous grove of quaking aspens known as Pando. Pando is, according to Wikipedia, "the world's largest tree ... A male clonal organism, Pando has an estimated 47,000 stems that appear as individual trees, but are connected by a root system that spans 106 acres". If I dig one of these stems up and plant it in my garden am I the proud owner of a tiny fraction of a tree in Utah, or has it now become a new organism?

I'll return to the question of defining organisms, but first I want to move to a more abstract level, metaphysics. Assume I am confident that the elephant in my room is an organism. What kind of an entity is an organism? The traditional philosophical answer, dating from Aristotle, is a substance. Unfortunately this is a technical term with several different meanings, but a rough more familiar equivalent is a thing —the word I shall generally use. What is a thing, in this philosophical sense?

All things but the very smallest are made of other things, ultimately molecules and atoms, themselves made of elementary particles. A thing typically has clear boundaries, and its existence is not dependent on other things; it is autonomous. Things do, of course, sometimes change, so one question that has been important in the metaphysical theory of things is what changes are consistent with the continued existence of a thing. The general answer to this, which again stems from Aristotle is that things have properties that divide between the essential and the accidental. Accidental properties may change, but for the thing to persist it must retain its essential property or properties. The essential property is what makes the thing the thing it is, in John Locke's words, "the being of anything whereby it is what it is".[3]

Note that while I have been talking about individual essences, not the essences of a kind that I criticised extensively in my earlier work, these are closely connected. In fact what makes the thing the thing it is is generally equated with what makes it the kind of thing it is. For the individual essence of Dumbo is generally assumed to be the property of being an elephant, which is to say, the possession of the essential property of elephants. As discussed in my last lecture, I do not believe there are any such essences of biological kinds, so here is a problem for the substantialist, or thing theorist. A natural response might be to appeal to the (often) unique genetic sequence of the individual. There are many prob-

[3] Locke (1689/1975), bk. III, chap. 3, sec. 15, p. 417.

lems with this idea, which I won't go into detail on here.[4] I will only note the curious consequence that if you have a monozygotic (i.e. identical) twin, you only become a separate entity from the twin when your genome first mutates in a unique way. Only through change, therefore, do you become an individual. But other problems with the substantialist view of the organism are clearer, and to these I now turn.

While an organism may look stable, autonomous and sharp-boundaried, we now know well that none of this is the case. First, we know that the stable form of an organism, a dog or an oak tree, is not just a default condition, but is maintained by trillions of chemical and physiological events happening every second inside it, the processes we call metabolism. I mentioned above that I like to define a process as an entity requiring change for it to persist. This is obviously true for a chaotic process such as Brownian motion, which just is movement, but it is equally true for the dog or the tree. Though the changes that sustain it may not be obvious, they are certainly real. An unchanging dog is a dead dog —though in truth a dog does not cease to undergo change when it dies, it merely moves from life-sustaining activity to the processes of decay.[5]

Second, organisms have life cycles. To take one familiar example, the life cycle of a beetle contains periods of time as an egg, a larva, devoted to eating, a pupa only metamorphosing, and an imago, or adult, mainly concerned with mating. These stages of the life cycle take on drastically different structures. Is a beetle egg a beetle? As a semantic question I assume little turns on this. But what is surely true is that it is the same organism as the beetle it later becomes (if it's lucky). If a thing is distinguished as a particular structure of its more fundamental elements, as traditional substantialism assumes, it is very hard to believe that an egg and a beetle could be the same *thing*. But of course they are stages in the same process.

Finally, and I think most importantly, organisms are not autonomous because, as we have increasingly come to understand, symbiosis is everywhere. We have all heard that 50% or more of the cells in our bodies are bacterial; and we are increasingly told of the benefits that we gain from having the right symbiotic microbes. As we become increasingly aware of the extent to which such symbionts are essential for our well-being, several philosophers are questioning whether these should not be thought of as mere collaborators in our lives, but as real parts of us (e.g., Triviño and Suárez 2020). Harking back to a term generally attributed to the great champion of symbiosis, Lynn Margulis,[6] they suggest we are really holobionts, composite organisms deriving from many evolutionary lineages. In further support of such a view it is beginning to appear that even viruses, which outnumber the cells in our bodies by a factor of ten, far from being potential enemies held in check by our ever-vigilant immune systems, may play essential roles in the symbiotic whole (Dupré and Güttinger 2016). Some protect against genuinely pathological agents; others store DNA that may be needed by our bacterial symbionts; and, most importantly, they may regulate the numbers of these bacteria.

I don't want to enter here into the debates about whether the organism is the holobiont. The important point for present purposes is just the difficulty of seeing a clear answer to this question. If we are interested in the totality of processes that promote the stability and survival of the human, then it is hard to see what sense there could be in excluding these vital parts of the total system. On the other hand, there is a plausible argument that for evolutionary purposes we may need to consider the discrete lineage that constitutes human cells, and that may share little of their history with many of their symbiotic partners. Echoing the promiscuous realism that I advocated in my earlier career, I have more recently proposed a *promiscuous individualism*. This is the idea that there may be no unique optimal way of carving individuals from the deeply entangled processes of life; only the best way of doing so for specific purposes.

This reflects a more general point about the entanglement of processes. Two entities interacting in such an intimate way that it is unclear whether there is one entity or two is a situation very hard to make sense of in a world of autono-

---

4 But see Dupré (2010).

5 One might wonder, I suppose, whether a dead dog was a dog. Happily, this is very much a problem for the substantialist, for whom the existence of a thing is expected to have sharp boundaries both spatially and temporally. For the processualist, such boundaries need be drawn only if and when they are required.

6 But incorrectly so, according to Baedke *et al.* (2020) who trace the term to the German theoretical biologist Adolf Meyer-Abich in the 1940s.

mous, hard-boundaried things. Think, for instance, of a relatively simple symbiotic system, such as a lichen. This is a functional whole composed of a fungus and a population of photosynthetic microbes.[7] Is this one thing, two things, or many things? Sometimes the interacting partners can survive on their own, sometimes not; but the particular mode of existence of the lichen is quite distinct from that of its constituents even when they have one. Individual processes, recall, are not typically autonomous, since dependent on interactions with their environments for their persistence, and seldom have sharp boundaries. There is no problem at all in thinking of the lichen as a living process within which constituent living processes may be identified as distinct individuals as we wish.

It should now be possible to see why a pluralistic metaphysics and epistemology fits so naturally with a process metaphysics. Pluralism is grounded in the observation that there is no unique way of dividing the world into kinds, and different projects of enquiry may call for different taxonomies. A simple way of understanding this is just to note that, at least in biology, every individual is different from every other and attention to different properties will distinguish different sets of items. The kinds distinguished are not unreal —hence the realism in my promiscuous realism— but they are not unique. Appreciation of the processual and entangled nature of biological individuals adds another dimension to this pluralism, what I have referred to as promiscuous individualism. The way we distinguish individuals from the complex and interconnected flow of biological process is also determined by the questions we are trying to answer or the goals we hope to pursue. If we are interested in the evolutionary history of a fungus that is sometimes part of a lichen, sometimes not, we will treat populations of fungal cells as the relevant units of analysis, only contingently connected to photosynthetic symbionts at particular times. But if we are interested in the physiology of the lichen, the fungus is merely a part of that whole.

We might think that this last example fits perfectly well with the now standard ontology of things and mechanisms. Things are composed of smaller things, and the interactions between them constitute mechanisms that explain the behaviour of the composite. A lichen, then, is a mechanism composed of fungal and microbial parts. In defence of such an interpretation, it is often remarked that scientists often think of themselves as looking for mechanistic explanations, and this perspective has been highly fruitful. The success of mechanistic thinking is at least something that the processualist had better give some account of.

Fortunately, this is easily done. Mechanism explains biological phenomena, conceived as the behaviour of an entity, as resulting from the behaviour and interactions of the parts of which that entity is composed. Both the entity referred to in the explanandum and the parts in the explanans are conceived of as things. The most famous modern statement of mechanism, the essay "Thinking about Mechanisms", by Peter Machamer, Lindley Darden and Carl Craver (2000) is explicitly ontologically dualist. What I have called "things" they refer to as "entities" and their behaviour and interactions are processes. I prefer to use the word "entity" as ontologically non-committal, but the point is clear enough. For me their entities are stabilised processes. The condition for their mechanistic explanations to be successful, in my view of the matter, is just that the things should be sufficiently stable for the time period over which they contribute to the mechanism (Dupré 2013). This is no doubt highly variable. If we are interested in the mechanism of continental drift, entities such as continents must be stable over many millennia. In the mechanisms of molecular biology sometimes milliseconds are sufficient.

If mechanism is translatable into process language, does the difference between these perspectives matter? Yes. It can matter very much if the theorist loses sight of the temporary nature of the stability of its objects and important dimensions of instability. Sabina Leonelli and I have given a detailed account of how just this treatment of very unstable processes as objects led to serious mistakes in the understanding and consequent policy relating to the Covid-19 pandemic (Dupré and Leonelli 2022).

Viruses, to begin with, are fascinating processes (Dupré and Güttinger 2016). It is tempting to identify the virus with the particles, virions, that we find on door handles and in the air we breathe, but this is quite wrong. The virus has a life

[7] Unsurprisingly, there are other symbionts involved in at least some lichens (Hodkinson *et al.* 2006), so the simplicity of the system should not be exaggerated.

cycle and the most interesting part of the life cycle is that which occurs after a cell has been infected. During that time, there is actually no individual entity that has a strong claim to be the virus; there is just a sequence of chemical activities in the cell. But more relevant to the pandemic is the epidemiology of a virus and, as we all came to understand during the pandemic, at this time scale viruses evolve rapidly. One problem to which this gave rise was with the naming of virus strains, which often gave people license to think that there was a name for a fixed kind of entity. In fact, this was never more than a phase in the evolution of a particular population; different names were given to relatively similar entities, and the same name was sometimes applied to very different entities.

A second problem concerned the scale at which we think of viral infection. If we were dealing with a homogeneous population of identical objects, this might not matter much. But the heterogeneity turns out to matter a lot. The severity of a viral infection appears often to turn precisely on the diversity of viral genomes, and the proper object of a particular infection is a diverse population, sometimes referred to as a quasispecies. More diverse populations can evolve more rapidly and thus more effectively evade the immune system. On the other hand if viral mutation is too rapid this may prevent the establishment of a stable infecting population, which has led to the investigation of mutagenic agents as possible therapies for viral infections.

All this raises a vital epistemological or methodological problem of maintaining the connection between changes in the phenomena and changes in the representations that attempt to track the phenomena. Both of these involve what we called reifications,[8] treating a phase of a process as a fixed entity, and reifications are essential for the scientific investigation of phenomena. But enabling the theoretical reifications to track the phenomenal reifications is a very difficult problem. Frequently the pandemic was described in terms that no longer adequately tracked the phenomena, and this led to problems in naming, as just mentioned, in testing, in monitoring infection rates, and in developing therapies and vaccines. In scientific management of such a rapidly changing process it is essential to develop methodologies that are firmly grounded in awareness of the dynamic character of the phenomena.

So far I have been concerned with arguing that process ontology better describes our scientific understanding and can lead to better science and science-based policy. But the implications extend beyond this. Humans are, of course, organisms, and I have been especially interested in recent years in exploring some of the implications of processualism for particular problems of human life, and it is to these that I now turn.[9]

3. Personal Identity

Let me start with a very traditional philosophical problem, that of personal identity. As most of you will know, this is the problem of understanding what it is for a person to persist over time. In what sense am I now the same person as the child who grew up in England in the middle of the last century? The problem is often set up with a problem such as John Locke's story of the prince and the cobbler (or shoemaker). These two people, according to the story, wake up one day each with all the memories and thoughts of the other. The story indicates the tension between two views of the problem. One that has much immediate appeal is that my continued identity is a mental matter; I have a continuous series of interconnected memories stretching into my distant past. On this view the prince now resides in the body of the cobbler, and vice versa. On a more materialistic view, both have undergone some bizarre psychological break.

The problem with the mentalistic view of personality is that it seems to assume a mind-body dualism that many now find incredible. The mind of the prince seems to be entirely detachable from his body and, therefore, presumably, something immaterial. Such a view lives on in the rather bizarre idea that we might perhaps survive forever in a computer

8 In Dupré and Leonelli (2022), we referred to the former as target reifications and the latter as means reifications.

9 The implications of seeing humans as processes are considered in more detail in Dupré 2025a.

programme, but for the more materialistically inclined, including myself, the mental is a feature of the organism, and not so easily detached. It is worth remarking that in the case Locke describes, a cobbler who wakes up and declares he is a prince is generally considered to have gone mad, and is probably in considerable danger for his life; safer to say he is Napoleon or Jesus Christ. The prince who has gone mad can probably expect more sympathetic treatment. It is important to recognise that it is only the simultaneity of these sudden onsets of insanity as the story is told that makes the mind exchange interpretation immediately plausible.

Increasingly over recent years, it has become more popular to accept the materialistic answer to the question, and personal identity has been associated with a spatio-temporally continuous body. A natural development of this idea has been to say that the identity of a person is just the identity over time of a particular (human) animal. This idea runs into problems, however, with the traditional view that an animal is a kind of thing. For the identity of a thing requires the continuity of certain properties, some of which are essential for that continuity. But the animal was once a zygote, a fertilised egg. I doubt whether I have many properties in common with a fertilized egg. The solution, of course, is to recognise that the animal is a process. Its continuity inheres in the causal connections between its temporal stages, not in any conserved essential property. Just as one stage of the organism *Rana temporaria*, the Common Frog, is a tadpole, but a tadpole is not a frog, so some part the life cycle of *Homo sapiens* is, or may be, a person.

This is important, because there are things we do, such as scientific research on embryos, that we would find quite inappropriate to do to a person. In the context of standard metaphysics this has led to some truly bizarre argumentation. The current policy in the UK is that it is permissible to experiment on human embryos until they are 14 days old. A common justification for this is that is the latest stage at which twinning can occur, when the embryo may divide into two parts, both of which may develop into human persons. If we considered the human life to have begun before this we have a potential contradiction: both twins would have been identical to the same embryo and, therefore, as identity is transitive (if A = B, and B = C, then A = C), to one another. But they aren't.

If the proposal that the point at which I began depends on an unrealised possibility seems odd, it is because it is. Fortunately, all of this trouble goes away when we move to a process ontology. Many processes divide: rivers, cells, species, political parties and much else. Identity of processes is certainly not transitive, or all life on earth would belong to the same species. How we name parts of a bifurcating process is a matter of convention rather than metaphysical fact. I am the same animal as my originating zygote, but no more the same person than I am the same adult: a zygote is neither a person nor an adult. The decision to designate part of the human life cycle as a person is a conventional decision based on whatever purposes we have in mind as attaching to that designation. And as I have already mentioned, distinguishing individual organisms from the flux of life is always partly conventional. The claim that Pando is one tree is exactly parallel to the argument that homozygotic, or identical, twins are one organism; thankfully, we have no overwhelming reason to adopt the latter —or for that matter the former— convention.

The process perspective, incidentally, has further interesting implications for the metaphysics of mammalian pregnancy.[10] This, it seems to me, is a paradigmatic instance of process bifurcation. There are interesting stages in this bifurcation that we may find relevant for various conceptual decisions. The process of a human life surely begins at fertilisation of the egg, though most such events do not lead a lot further. Implantation is important at least in marking a great increase in the probability of continuation of the process. Viability of the foetus is important in providing more options for the continuation of the life cycle, outside the mother's body. And birth is important in providing a massive reduction in the interdependence of the mother and child, though surely not a total one unless we equate interdependence with necessary interdependence. This perspective gives us much more tractable, though no doubt still difficult, approaches to questions about abortion, foetal and maternal rights and much else including embryo research.

[10] For a detailed elaboration of a processual view of pregnancy, see Meincke (2022).

4. Human development and classification

Another set of questions for which a process perspective is transformative are those arising in the often highly controversial classification of humans, a topic I discussed in the previous article. A deeper understanding of the difficulties faced by human classification comes from the recognition that humans are not things with essences but developmental processes, and developmental processes with huge sensitivity to both external and internal influences. This results in a great diversity of properties among adult humans. I say adult humans, because the diversity develops as the process continues.

A perhaps surprising thinker to recognise this point clearly, one of the most misunderstood of philosophers, was Adam Smith. Smith is, of course, famous for his emphasis on what is undoubtedly a crucial feature of the human species, the division of labour. Especially among right wing enthusiasts for his work, this is often connected with differences in innate ability, taken to justify differences in status and wealth. But as Smith writes in *The Wealth of Nations*:

> The difference of natural talents in different men, is … much less than we are aware of; and the very different genius which appears to distinguish men of different professions … is not … so much the cause, as the effect of the division of labour. The difference between … a philosopher and a common street porter, for example, seems to arise not so much from nature, as from habit, custom, and education. … [F]or the first six or eight years of their existence, they were, perhaps, very much alike … About that age … they come to be employed in very different occupations. The difference of talents … widens by degrees, till at last the vanity of the philosopher is willing to acknowledge scarce any resemblance. (Smith 1776, bk. 1, ch. 2, para 4.)

The point has very wide applications to human kinds. Generally we should look for an understanding of the differences between humans of different kinds in the diversity of developmental histories, much of which is due to different environmental conditions. This insight provides a deeper understanding of the complexities of sex and gender and of race that I discussed in the previous paper, and I shall now briefly return to these.

I mentioned in my last paper that we should see sex and gender classifications as the outcomes of a complex and multicausal developmental process, and I hope in this paper to have made the grounding of this claim clearer. The core biological distinction of production of gametes of different sizes, as well as other biological distinctions based on physiology, genes or hormones all generate their own, sometimes divergent classifications, and social factors add another level of complexity. The diversity of relevant kinds of people is well-captured by the ever-growing acronym, LGBTIQA+. I think it is helpful to follow feminist biologist Anne Fausto-Sterling (2012) in attributing to people a sex/gender rather than either a sex or a gender tout court. This leaves us open, in our more careful talk, to use strictly sex-referring terms, male and female, in the sense recommended by Paul Griffiths (2021), to refer to a property of some organisms at particular stages of their life cycles, as they produce gametes of a particular size. I say "our more careful talk"; I assume we can still use the words male and female to refer to the primary way most —but not all— people choose to present themselves, but the more technical usage reminds us that these terms group together a great variety of developmental outcomes. Any fine-grained classification of sex/gender is difficult, variable and imprecise. And we should, as I stressed in the previous paper, always be prepared to ask exactly why it is necessary or important to classify people at all.

A final look at the concept of race requires me to digress a little and introduce more explicitly another concept that is central to a processual biology, that of lineage. For race is not an essential property of individuals but, if anything, a group level characteristic of the human species as a whole, a division of the human lineage into sublineages. A lineage is an ancestral/descendant sequence of populations, connected to one another, and more or less separated from other populations. It may be a very long-lasting process, stabilised by natural selection but also by sexual reproduction, which connects the individual members of a lineage and enforces boundaries between lineages, and in some cases by cooperation between the organisms that make up the populations. Lineages are now the vital context in which most evolu-

tion takes place. They are not, however, essential for evolution, at least in the form we now know them; indeed they are themselves evolved and evolving processes. Sexual reproduction, which provides the boundaries for many of the most successful currently existing lineages, is something that evolved. Similarly cooperation, the feature that does most to explain the great success of our own species, is a variable feature of lineages, largely absent from many or most.[11]

Lineages, I have said, are processes —like organisms, structures more or less stabilised within the flow of life. Philosophers of biology will be reminded of the once shocking and controversial claim by David Hull and evolutionary biologist Michael Ghiselin, that species, assumed to be lineages, branches of the tree of life, were individuals. The shocking nature of this claim, I suggest, derived from the assumption that this consigned species to the category of thing, and in many ways this seemed bizarre.

But the realisation that they are processes that can persist over long periods of time, removes much of this concern. It also has great positive benefits. It is no surprise that, as processes, lineages should lack clear boundaries and require pragmatic and partly conventional decisions as to how they are distinguished within the general flow of biological reproduction and change. We now know that hybridisation is extremely common between related species, and all kinds of temporarily more or less isolated currents can exist within them. Debates between so-called lumpers and splitters on the correct fineness of grain of our classifications are generally questions of pragmatics rather than debates about the nature of reality.[12]

And so finally back to race. The human species is very numerous and widely spread geographically. Though genetically remarkably homogeneous, it is phenotypically and culturally diverse. There are barriers within it such as oceans and mountain ranges, as well as distinct cultures, more or less impermeable to one another. Inevitably this implies many transitory subcurrents within the overall evolutionary —broadly construed to include cultural as well as biological factors— trajectory of the whole. As noted in the preceding paper, some superficial features, notably skin colour, change rapidly in response to environmental conditions, and such features can provide easily applied markers for temporary human currents. But crucially, it is also a species that is genetically surprisingly homogeneous, and highly migratory. Migrating humans cross oceans and climb mountains. Human groups adopt cultural innovations from other groups with which they interact.

So within the turbulent flow of the human lineage it is possible, if we wish, to pick out countless temporary subcurrents. We distinguish one another by religion, nationality, language, cultural and artistic tradition, and much else besides as well as race. These subcurrents, providing a vast diversity of technologies, arts, philosophies and much else, offer huge potential benefits to the species as a whole. But, as I emphasised in the previous paper, we must always ask of such distinctions what purpose they serve. The trouble with race as a category for distinction, is that these purposes have been almost uniformly malign, an excuse for the exploitation of and expropriation from vast numbers of people. Seeing races as no more than temporary eddies within the flow of human change undermines any possible biological justification for these ills. And, as I noted in the last paper, the apparent stability of these eddies is largely explained precisely by the malign purposes to which they are put, racism.

It is sometimes assumed that this kind of denial of sharp boundaries between kinds is a kind of antirealism about distinctions. But the point is not that there are no boundaries, but there are many, and these are complexly nested and cross-cutting.[13] We can find real distinctions between kinds of people, but these are not distinctions between different kinds of thing, demarcated by their unique essences, but one of many ways of dividing people into different categories for particular purposes. The real question to ask about both sex/gender and race distinctions is what exactly these purposes are. Often the answer to this question is distinctly unedifying.

---

11 Cooperation, together with massive diversity of developmental trajectories, is what makes possible the division of labour that, in turn, enables the vastly complex societies in which we live. I would love to say more here about the absurdity of the frequent attempts to ground in biology the competitive individualism that has so disastrously dominated so much recent social and political thought, but limits of space dictate that I leave that for another time and place. The topic is discussed, still too briefly, in Dupré 2025a.

12 Lineages as processes are discussed in detail in Dupré 2025a, chap. 2.

13 The "realism" in "promiscuous realism" is intended to be taken seriously.

5. *Free Will*

Let me mention one final topic to which I think the metaphysical perspective I advocate is relevant, the freedom of the will. Philosophical orthodoxy sees humans as complex machines, parts of a deterministic world. Compatibilists reassure us that we still have as much freedom as we have any reason to expect or want. Many are understandably unconvinced, but if the alternatives are either to see ourselves as unique exceptions to an otherwise seamless web of causality, or as random action generators, we are probably better off with the consolations of deterministic compatibilism.

If, however, as I argue, the world is one of open-ended processes providing pockets of partial order in a generally disordered world, a robust concept of freedom does not require that we see ourselves as exceptions of kind to the rest of the natural order. Freedom as autonomous self-direction remains a tricky concept to articulate properly, but it is not hopeless in the way that it is in the context of the deterministic billiard ball world.

I can only summarise the way this goes in the present paper.[14] Much discussion of human action still begins with the assumption that freedom requires no more than that we act to pursue what we want most in the light of what we believe will best achieve those ends. This is how the machine ideally works; unfreedom comes when we are prevented from acting in this personal utility-maximising way by external interventions of various kinds.

I object that when I lie in bed knowing that I need to work on my Lullius lectures, it may well be that what I want to do most is to go back to sleep, and I believe I can do this if I turn off the alarm. Nonetheless, I get out of bed and get to work. Freedom, I suggest, is precisely the ability *not* to do what I want to do most at the moment, but rather to pursue longer term goals that ultimately shape the world around me in ways that I choose. This is not a particularly novel perspective; it is perhaps a very soft version of Kant's austere vision of freedom as duty. What the metaphysics of disorder and process contributes is just to deprive such a story of philosophical mystery. The disordered world offers many affordances of many different kinds for the partial creation of order, and plans, principles or long term goals are all ways of engaging such affordances.

I must add that the question can, of course, be pushed further back. Commitments and principles of action allow me to —no doubt very slightly— shape the world rather than merely allowing it to shape me. But where do my commitments and principles come from? Much that shapes the process that I am —my upbringing, education and culture, for example— is largely a matter of luck. Perhaps there is some individual reflection on who I want to be for which I can claim more credit. But for now, I will only say that freedom, like explanation, has to stop somewhere. We are not self-creating pure wills, but parts of a complex and deeply interconnected lineage. The possibilities for autonomous action within that bigger context is really all the freedom that is worth having.

6. *Conclusion*

I end these papers (Dupré 2025b and this) with two very general reflections. First, I have heard friends and colleagues express amazement that, having spent half my philosophical career defending pluralism against reductive monism —there are only physical particles and structures of those— I now seem to have embraced a radical metaphysical monism: there is only process. I do not see this as particularly strange. I am a pluralist about the structures that emerge in our world, with their novel and distinctive properties and capacities. But metaphysically there is surely only one world. Of course, one world might contain both substances and processes, and perhaps other entities besides. But this is a quite different question from the pluralism at the level of the empirical world. Moreover, and this is the main point, the metaphysical monism of process precisely explains the empirical pluralism. That is something I have tried to show in the present work.

---

[14] The argument is presented in more detail in Dupré 2025a, ch. 6.

Finally, is there anything to be said about why the world is, if it is, the kind of world I have tried to describe. I have remarked that cosmologists now largely agree that the universe began with complete chaos. Gradually, over billions of years and across vastly different time scales, pockets of order emerged, from inconceivably minute particles to immense galaxies. But why was the original chaos disposed to settle, here and there, into these stable forms.

David Hume offered part of the answer in a lovely passage from the *Dialogues concerning Natural Religion*:

> Suppose ... that matter were thrown into any position, by a blind, unguided force; ... this first position must, in all probability, be the most confused and most disorderly imaginable ... [S]uppose that the actuating force, whatever it be, still continues in matter, this first position will immediately give place to a second, which will likewise in all probability be as disorderly as the first, and so on through many successions of changes and revolutions. No particular order or position ever continues a moment unaltered. ... Every possible situation is produced, and instantly destroyed. ...
>
> Thus the universe goes on for many ages in a continued succession of chaos and disorder. But is it not possible that it may settle at last ... so as to preserve an uniformity of appearance, amidst the continual motion and fluctuation of its parts? This we find to be the case with the universe at present. Every individual is perpetually changing, and every part of every individual; and yet the whole remains, in appearance, the same. ... [T]his adjustment, if attained by matter of a seeming stability in the forms, with a real and perpetual revolution or motion of parts, affords a plausible, if not a true solution of the difficulty. [Hume (1779/1947), pt. 8, paras. 8-9.]

This survival of the stable is the powerful tautology at the heart of Darwin's theory of natural selection, articulated into a detailed empirical theory by almost two centuries of subsequent scientific work. And surely a similar account can be given of the earlier emergence of physical structures that cosmologists and theoretical physicists have described more recently.

There is, of course, still a further question we may want to ask. Why should this matter first thrown into its "confused and most disorderly imaginable" position have the capacity to form itself into the exquisite structures we see today? Why for instance should the particles of matter attract one another to form the stars and planets our universe contains? Far more remarkably, why should the particles have the ability to come together to produce objects with the extraordinary properties of water or of carbon atoms that seem so remarkably designed to contribute to living metabolism? This, I think, is the last resting place for natural theology: if some omnipotent being had lit the fuse that set off the big bang, surely She might have imbued its matter with just such properties as to make the emergence of complex structure possible.

Sadly or otherwise, as Hume so powerfully argues in the Dialogues, this is a woefully inadequate abductive argument for a creator, let alone any particular kind of creator. Philo, the character most closely reflecting Hume himself in the Dialogues, has no trouble rattling off equally improbable hypotheses. This really is where, at least for now, explanation has to stop or, in Hume's words: "A total suspense of judgement is here our only reasonable resource".

Acknowledgements

This and the previous paper ("Disorder", Dupré 2025b) are based on the Raimundus Lullius lectures given in Oviedo, Spain in July 2024 at the 11th annual meeting of Spanish Society for Logic and Philosophy of. Science. I am very grateful to the organisers of the conference, Cristina Corredor and Javier Suárez, first for inviting me to give the lectures, and then for making my visit to Oviedo so memorably enjoyable.

REFERENCES

Baedke, J., Fábregas-Tejeda, A., and Nieves Delgado, A. (2020). The holobiont concept before Margulis. *Journal of experimental zoology Part B: Molecular and developmental evolution, 334*(3), 149-155.

Dupré, J. (2008). *The constituents of life*. (The Spinoza Lectures, University of Amsterdam, 2006). Van Gorcum. Reprinted in Dupré (2012), chs. 4 and 5.

Dupré, J. (2010). The polygenomic organism. In S. Parry and J. Dupré (Eds.), *Nature after the genome* (pp. 19-31). Blackwell. Sociological Review Monograph series. Reprinted in Dupré (2012).

Dupré, J. (2012). *Processes of life: Essays in the philosophy of biology*. Oxford University Press.

Dupré, J. (2013). Living causes. *Proceedings of the Aristotelian Society, supplementary volume, 87*, 19-38.

Dupré, J. (2025a). *Everyone flows: A process perspective on human life*. Oxford University Press.

Dupré, J. (2025b). Disorder. *Theoria. An International Journal for Theory, History and Foundations of Science*, 40(1), 5-16.

Dupré, J., and Guttinger, S. (2016). Viruses as living processes. *Studies in history and philosophy of science Part C: Studies in history and philosophy of biological and biomedical sciences, 59*, 109-116.

Dupré, J., and Leonelli, S. (2022). Process epistemology in the COVID-19 era: Rethinking the research process to avoid dangerous forms of reification. *European journal for philosophy of science, 12*(1), 20.

Fausto-Sterling, A. (2012). *Sex/gender: Biology in a social world*. Routledge.

Griffiths, P.E. (2021). What are biological sexes? *PhilSci Archive*. https://philsci-archive.pitt.edu/19906/1/Griffiths%20-%20What%20are%20biological%20sexes.pdf.

Hodkinson, B. P., Lutzoni, F. M., Loveless, T. M., and Bishop, P. E. (2006). Non-photosynthetic bacteria and the lichen symbiosis. In Bright, M., Horn, M., Zook, D., Lücker, S., and Kolar, I. (Eds.), *Fifth international symbiosis society congress: program, abstracts, participants* (p. 15). Promare.

Hume, D. (1779/1947). *Dialogues concerning natural religion*. Macmillan.

Locke, J. (1689/1975). *An essay concerning human understanding*. Oxford University Press.

Machamer, P., Darden, L., and Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, *67*(1), 1-25.

Meincke, A. S. (2022). One or two? A process view of pregnancy' *Philosophical studies*, *179*(5), 1495-1521.

Nicholson, D. J. and Dupré, J. (eds.) (2018). *Everything flows: Towards a processual philosophy of biology*. Oxford University Press.

Priani, E. (2021). Ramon Llull. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). URL = <https://plato.stanford.edu/archives/spr2021/entries/llull/>.

Smith, A. (1776/1937). *An Inquiry into the nature and causes of the wealth of nations*. Modern Library.

Triviño, V., and Suárez, J. (2020). Holobionts: Ecological communities, hybrids, or biological individuals? A metaphysical perspective on multispecies systems. *Studies in history and philosophy of science Part C: Studies in history and philosophy of biological and biomedical sciences, 84*, 1-11.

John Dupré is Professor of Philosophy of Science at the University of Exeter (UK), where he directed Egenis, the Centre for the Study of Life Sciences, for 20 years (2002-2022). He specializes in the philosophy of science, with a particular interest in the philosophy of biology.

Address: Egenis, Centre for the Study of Life Sciences; Department of Social and Political Sciences, Philosophy, and Anthropology; University of Exeter; Byrne House; St. German's Road; Exeter EX4 4PJ. Email: j.a.dupre@exeter.ac.uk – ORCID: https://orcid.org/0000-0002-7451-2127

Articles

Theoria, 2025, 40(1), 30-46
https://doi.org/10.1387/theoria.26849

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

The harms of non-derogatory uses of slurs and the Potential Normalization Argument

(Los daños de los usos no derogatorios de los slurs *y el Argumento de la Normalización Potencial)*

Dan Zeman*
University of Porto
https://orcid.org/0000-0001-5620-0042

Keywords

Slurs
Derogatory and non-derogatory uses
Harm
Normalization

ABSTRACT: There is no doubt that slurs harm. They do so by denigrating their targets, by putting them down, by marginalizing them. This is why in many legislations around the world, the use of slurs has been banned or penalized. But should *all* uses of slurs be banned? Many uses of slurs seem to be non-derogatory and to have beneficial effects. However, such uses are double-faceted: as both armchair reflection and experimental studies have shown, they are able to produce harm as well. In this paper, I approach the broad question of whether all non-derogatory uses of slurs should be banned. I first present the main uses of slurs that have been considered to be non-derogatory and recent reactions to those. The upshot of this survey is that uses of slurs that have been considered non-derogatory do, in fact, produce harm. I also flag what various authors have recommended in relation to the issue of banning such uses. Against this background, I engage with a recent view put forward by Alba Moreno Zurita and Eduardo Pérez-Navarro, who urge extreme caution with respect to any uses of slurs, due to their potential to normalize derogation. After presenting their view and their main argument, I raise an objection related to their treatment of neutral uses of slurs. I end with pointing out that, while their endeavour has merit in that it pushes the discussion further, it raises certain issues —of both an empirical and a normative nature— that need to be addressed.

Palabras clave

Slurs
Usos derogatorios y no derogatorios
Daño
Normalización

RESUMEN: No cabe duda de que los *slurs* hacen daño. Lo hacen denigrando a sus destinatarios, menospreciándolos, marginándolos. Por eso, el uso de *slurs* ha sido prohibido o penalizado en muchas legislaciones de todo el mundo. Pero ¿deberían prohibirse todos los *slurs*? Muchos de sus usos parecen no ser derogatorios y tener efectos beneficiosos. Sin embargo, tales usos tienen dos caras: como han demostrado tanto la reflexión de sillón como los estudios experimentales, también tienen la capacidad de producir daño. En este artículo me planteo la cuestión general de si deberían prohibirse todos los usos no derogatorios de los *slurs*. Primero, presento los principales usos de los *slurs* que se han considerado no derogatorios y las principales reacciones recientes a los mismos. El resultado de este estudio es que los usos de los *slurs* que se han considerado no derogatorios, de hecho, producen daño. También señalo lo que varios autores han recomendado en relación con la cuestión de prohibir tales usos. En concreto, me ocupo de una posición reciente presentada por Alba Moreno Zurita y Eduardo Pérez Navarro, quienes instan a extremar la precaución respecto a cualquier uso de los *slurs*, debido a su potencial para normalizar la derogación. Tras presentar su punto de vista y su argumento principal, planteo una objeción relacionada con el tratamiento que dan a los usos neutrales de los *slurs*. Concluyo señalando que, si bien su esfuerzo es meritorio en la medida en que hace avanzar el debate, plantea ciertas cuestiones —de naturaleza tanto empírica como normativa— que es necesario abordar.

* ***Correspondence to:*** Dan Zeman. Institute of Philosophy, Faculty of Letters, University of Porto, Via Panorâmica, s/n, 4150-564, Porto, Portugal – zeman@letras.up.pt – https://orcid.org/0000-0001-5620-0042

How to cite: Zeman, Dan (2025). «The harms of non-derogatory uses of slurs and the Potential Normalization Argument»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 30-46. (https://doi.org/10.1387/theoria.26849).

Received: 07 August, 2024; Final version: 24 April, 2025.

ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

This work is licensed under a *Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International* License

There is no doubt that slurs harm. They do so by denigrating their targets, by putting them down, by marginalizing them. Worse still, slurs can be an aid to dehumanization and sometimes they can play a crucial role in bringing about violent outcomes such as genocide. Additionally, experimental studies in psychology have conclusively shown that slurs affect their targets and lead to a decrease in their well-being, by influencing their self-worth and performances, and increasing stereotype- and identity-threat. This is why in many legislations around the world, slurs (or at least their derogatory uses) have been banned or penalized.

But should *all* uses[1] of slurs be banned? First, at least on the face of it, it seems that using slurs in some contexts can be beneficial. For example, when a court is investigating, say, a case involving verbal aggression, it might be crucial for the jury to know exactly what has been said in order to reach the right verdict. Similarly, when journalists report a story involving the proffering of slurs, it might be important to reproduce the same words to make the information as exhaustive as possible in order to arrive at a maximally detailed description of the facts. Finally, in research and teaching, those involved in these activities often use examples to introduce their topic of study (slurs, pejoratives, hate speech, etc.), to get a certain point across, or to support an argument. In all these cases, slurs are being used (although in quotes, or by using other detachment devices). Is it reasonable to ban such uses?

Second, and perhaps more substantially, many discriminated communities have used slurs to overcome precisely the discrimination they have been historically subjected to, appropriating or reclaiming the word that has been used as a weapon against them. Changing a slur by endowing it with positive meaning instead of a negative one has been a powerful tool in this process. Furthermore, as several authors have argued, slurs can also be employed neutrally —as when a slur is used to refer to someone's friends, or as a mere tag to self-identify as belonging to a certain group (an ethnic one, for example). Should *these* uses be banned?

Arguably, the uses of slurs listed above are non-derogatory —or at least have been considered so. While the case for banning derogatory uses of slurs across the board is relatively easy to make (due to their obvious pernicious effects, or the harm they produce[2]), the same cannot be said about non-derogatory uses. The issue is complicated, because non-derogatory uses of slurs seem to be double-faceted: on the one hand, as we just saw, such uses of slurs can be beneficial, and depending on the type of use, they can play a more or less important role in a community; on the other hand, it has been argued (as we will see shortly) that some of what are taken to be non-derogatory uses of slurs can have pernicious effects, too. So, the question becomes, in what contexts and under which circumstances should purported non-derogatory uses of slurs be banned? This is the broad issue I address in this paper.

To approach it, I will first expand on what were thought to be, at different points in time and by various authors, non-derogatory (and thus non-harmful) uses of slurs, together with more recent reactions to such uses (section 1). What

---

1 I'm employing here the term "use" to refer to any occurrence of a slur (either oral or in writing), regardless of the linguistic embedding or non-linguistic context. One consequence of this usage is that the familiar "use- mention" distinction that philosophers of language often warn about collapses; for me, mentioning a slur is one type of use of that slur, regardless of the exact mechanism by which mentioning is communicated (quotes, air quotes, tone, implicit understanding, etc.). It is perhaps worth noting that, despite its treatment being a contentious issue, this take on the use-mention distinction is consistent with certain theoretical approaches to quotation (e.g., Frege's *Identity Theory*; see García-Carpintero, 2004, p. 678). I embrace this broad sense of "use" because although I will talk about both cases of mentioning slurs and of slurs being used in non-derogatory ways without being mentioned, I'm interested in the question whether *all* non-derogatory uses of slurs should be banned.

2 I'm operating in this paper with a very loose notion of harm, one that comprises any (systematic, targeted and interpersonal) negative effect a slur can have on a member of the target group. This opens up at least a couple of important issues: whether any such harm should be taken into consideration when banning a certain word is discussed and whether the harm produced can be considered to be the direct result of the semantic profile of a slur or of other, perhaps pragmatic or simply sociological, factors. I take here the stance (without arguing for it) of allowing any (systematic, targeted and interpersonal) harm to be at least a prima facie good reason to consider banning a certain word. As for the second issue, I remain neutral on the issue of whether the relevant harm is necessarily a result of the semantic profile of a slur or not, as I don't think anything I will say in what follows trades on taking one view or another. All I assume is that there is a connection between harm and the derogatory character of a slur, regardless of how this is spelled out theoretically.

emerges from considering these reactions is that non-derogatory uses of slurs can also have pernicious effects and thus cause harm. After flagging what the respective authors recommend vis-à-vis various uses of slurs previously considered to be non-derogatory, I move on to the main focus of this paper: a recent, more general, argument found in the work of Alba Moreno Zurita and Eduardo Pérez-Navarro who argue that *any* use of a slur is potentially harmful, and thus urge caution when using slurs. I present their main argument (which I reconstruct and dub "the Potential Normalization Argument") and other important claims they make in section 2, while in section 3 I bring forth an objection to their view. I take the argument to be an important one to discuss, regardless of whether the objection is successful. Thus, while this paper is quite limited in scope, I take it that engaging with their view pushes the discussion in new, interesting directions.

1. The uses of slurs and their harms

While it is widely agreed that the main function of slurs is "to derogate or dehumanize (...), to signal that their targets are unworthy of equal standing or full respect as persons" (Jeshion, 2013, p. 232), they have many uses, not all of them derogatory. As a broad two-partite categorization, one can group uses of slurs in *derogatory* (also called "weapon uses" by Jeshion, 2013) and *non-derogatory* ("non-weapon uses"). In what follows, I will take each broad category in turn, underlying some of the ways they have been considered to produce harm. In connection to non-derogatory uses, I will showcase several such presumed uses, and flag what various authors reacting to them have recommended vis-à-vis their banning.

1.1. Derogatory uses and their harms

Derogatory uses are the closest to fulfilling the main function of slurs spelled out above, and they are illustrated by sentences like "John is a boche."[3], "You are a boche.", "You, boche!" and so on.[4] The pernicious effect of derogatory uses of slurs is undeniable, and it has been amply documented, with the types of harms they produce varying from severe forms to less so.

Thus, as work on dehumanization shows (see chiefly Livingstone Smith, 2011, 2020, 2022), slurs have been one of the tools by which people dehumanize those they deem inferior. Dehumanizing one's (perceived) enemies makes it easier to isolate, discriminate, and marginalize them, as not being on the same level on the Great Chain of Being. Further, slurs are an effective way to help bringing about violent outcomes when parts of societies clash: as the widely cited work by Tyrrell (2012) concerning the Rwandan genocide shows, the use of slurs and other words such as "cockroaches" and "snakes" has been a crucial aid to starting and justifying the massacre of the Tutsi and of other minorities by the Hutu. Less dramatically, but not less importantly, experimental studies in psychology have shown that being targets of slurs leads to internalization of negative attitudes and associations (Carnaghi & Maass, 2008), to feelings of alienation due to exclusion and to a "us vs. them" frame of mind, which in turn lead to a decrease in self-assuredness and self-esteem (Carnaghi & Maass, 2008), and to stereotype threat, which leads to worse performances by the members of the target group due to worries about conforming to the stereotype (Steele *et al.*, 2002).[5] As can be seen from this list, while not all uses of slurs end up having dramatic effects such as dehumanization or genocide, even the most mundane derogatory use of a slur can affect its target in significant ways.

---

[3] Whether to use slurs (in quotation marks) in academic papers is one of the issues this paper deals with. As can be seen, here I take the widespread position that mentioning some slurs is important.

[4] There are nuances about the different ways such constructions derogate: for example, *calling* someone a boche instead of merely *asserting* that they are a boche might have a slightly different derogatory profile. Interesting as they are, I will leave such issues aside.

[5] I'm following here the very useful review of the experimental literature in (Herbert, 2017).

It is generally contended that at least part of the harm produced by slurs comes from normalizing certain ways of thinking and talking about the target group, which makes it easy to act on the beliefs that possibly lead to the outcomes described in the previous paragraph. In this, slurs have been thought to be intimately connected with pernicious ideologies, with the relation between them being one of *mutual reinforcement*: the derogatory use of slurs is made possible by the ideologies that consider members of the target group to be inferior as humans, while repeated use further strengthen those ideologies, perpetuating negative stereotypes about the members of the target group and essentializing them.[6] It is also important to note that slurs do so often implicitly, escaping people's conscious efforts and relying instead on their biases (Herbert, 2017).

1.2. Non-derogatory uses and their harms

Non-derogatory uses of slurs come in many forms.[7] Below I present several types of uses of slurs that have been taken to be non-derogatory, together with more recent reactions to them. Subsection 1.3. discusses the upshot for the issue tackled in this paper.

1.2.1. Negation

Denying that a slur applies to someone by means of sentences like "Hans is not a boche." or "There are no boches; there are only Germans." has been considered a non-derogatory use of the slur. Thus, such sentences have been taken as "disowning the [derogatory] attitude" (Blackburn, 1984, p. 148), to "convey that ['boche'] is not something one calls anyone" (Hornsby, 2001, p. 129) or to be a way of "repudiating the use of the term" (Dummett, 2007, p. 527). Other authors take the negation in such sentences to be metalinguistic (e.g., Hom, 2008).

However, despite assurance from these philosophers, it has been remarked that denying that a slur applies to someone by using negation (as in the sentence "Hans is not a boche.") is not the most efficient way to go when one intends to defuse the effect of a slur, as its presence, even within negation, manages to offend (the speaker is using the slur, even if their intentions are to defuse it). Anderson & Lepore (2013) thus claim that "a denial of ['Hans is a boche.'] is no less inflammatory than ['Hans is a boche.'] itself and don't recommend this strategy when attempting to reject a derogatory, weapon-use of a slur" (p. 28).

1.2.2. Speech and belief reports

Disquoting a sentence in an indirect speech or a belief report is usually a correct way to convey what the reported person said or believed. Thus, reporting utterances containing a slur has been considered non-derogatory, the embedding of the slur under the speech or attitude verb insulating its derogatory character (in the sense that it is not the speaker who derogates, but the person whose speech or belief is reported). Schlenker (2003) has provided some examples purporting to show the insulating effect of such reports, the most well-known being "I am not prejudiced against Caucasians. But John, who is, thinks/claims that you are the worst honky he knows" (Schlenker, 2003, p. 43).

As with negation, it has been recently contended that what has been claimed to be an insulating effect of speech and attitude verbs is, in fact, a defective environment (with Schlenker's example being heavily contested). What many authors

---

6 For concrete proposals about how the two relate, see Hom (2008), Kukla (2018), Popa-Wyatt & Wyatt (2018), Davis & McCready (2020), Swanson (forthcoming) —among many others.

7 For a list, partially repeated here, see Cepollaro & Zeman (2020). That list also contained fictional uses of slurs (both as in real world slurs used in fiction —e.g., the n-word used in movies, and as slurs created for fictional characters— e.g., the slur "toaster" used for replicants in the series *Battlestar Galactica*). Slurs used in humour or for comedic effect can also be counted among non-derogatory uses. Although they are important and raise a variety of issues, I will ignore them in what follows. See, however, Anderson (2015) for an introductory discussion of the topic.

agree about is that, in reporting a use of a slur, the speaker cannot simply claim innocence, since the slur is being used in the report as well and derogation is traced to the speaker. Additionally, there is ample recent experimental evidence that reported slurs, while less derogatory than non-embedded ones, are nevertheless perceived as offensive (Cepollaro *et al.*, 2019; Cepollaro *et al.*, 2024). In fact, it is nowadays agreed that slurs "scope-out" from any kind of linguistic environments —including conditionals like "If Hans is a boche, we shouldn't invite him for dinner." This characteristic of slurs has been dubbed "hyper-projectivity" or "non displaceability" (Potts, 2007; Hom, 2010; Croom, 2011; Cepollaro, 2020; etc.) and refers to the fact that the derogatory character of slurs doesn't disappear even when they are embedded in linguistic constructions. As the examples discussed in these paragraphs show, despite slurs appearing embedded under negation, indirect speech and belief reports, and conditionals, they retain their derogatory character —just as they do when used unembedded.

1.2.3. Quotation

Many uses of slurs are didactical or quotational —such as those already mentioned at the beginning of this paper. Quotation, the widespread device by which one insulates the meaning of words to focus on the words themselves, and which is the predominant tool in court, newspapers and classrooms, has once been thought to provide a safe haven for slurs. Thus, Hom takes "explicit quotation in the courtroom" to be among acceptable uses of slurs (2008, p. 427), Williamson believes that "mere quotation marks (...) isolate us from [a slur's] derogatory implications" (2009, p. 139), while Hornsby claims that "quotation has some sealing off effect" (2001, p. 130).

Despite widespread past consensus that quotation isolates the derogatory character of slurs, it has, again, been noted that quoted slurs can derogate as well. Famously, and in opposition to (the then) orthodoxy, Anderson & Lepore note that the occurrence of a slur in a sentence like "'Boche' is a term for Germans" "can easily cause alarm and offence" (2013, p. 37) and counter the more lenient claims of Hom, Williamson and Hornsby by noting that, even in such contexts, there is a preference for using more acceptable formulations (e.g., "the n- word") instead of the slur (in this case, the n-word). Thus, in relation to such uses, they propose the following strategy: "As a safeguard against such inurnment, we strongly urge you always to ask yourself how a targeted member, perhaps accidentally overhearing you, would react to your usage. You'll find, as we have, that much of what seems suitable is definitely not" (Anderson & Lepore, 2013, p. 31).

This noted tendency of slurs' derogatory aspect to "scope-out" of quotation and produce harm is supported by empirical research. First, in the same study mentioned above in relation to speech reports (Cepollaro et al, 2024), the authors have found that slurs reported both in direct speech ("Z said that 'Y is an S'") and in mixed quotation ("Z said that Y is 'an S'"), while less derogatory than non-embedded ones, are still perceived as offensive. Second, in some of the experimental studies surveyed by Herbert (2017), participants were subjected to single printed slurs (Carnaghi & Maass, 2008) or received subliminal exposure of slurs (Fasoli *et al.*, 2016), and a similar effect has been found to participants being subjected to uses of slurs in a context. Herbert takes this to show that "both using and mentioning the slur can prime audience members in this harmful way" and that "[m]erely mentioning a slur can act as a powerful priming mechanism of pernicious implicit associations" (2017, p. 144). While this result has been achieved in testing outgroup members, it is reasonable to claim that it also holds for members of the target group. To put it succinctly, the experimental evidence she samples shows that there is not much difference between uses of slurs in context and mere occurrences of slurs in terms of the harm produced; in her view, this "raises important ethical questions for how we ought to talk about slurs" (Herbert, 2017, p. 147).

1.2.4. Appropriation

A great number of slurs have been appropriated (reappropriated, reclaimed —the terminology varies; I will use these terms here interchangeably) or are in the process of being so. For example, "queer" is often cited as a successful instance of an appropriated slur (as the titles of journals like *Queer Studies* or of TV shows like *Queer Eye* attest), as well as the n-word, when used within certain communities in the U.S.A. Appropriation itself has many facets, and scholars have distinguished at

least two types: i) appropriation within political activism; ii) appropriation as signalling or fostering camaraderie, with many other uses being attested too (Naylor, 1986; Kennedy, 2003; Jeshion, 2013, 2020; Bianchi, 2014; etc.).

What about these uses? While generally the use of slurs within the process of appropriation has been looked upon favourably, there have been voices pointing out that reclamation projects have their own risks. First, there are various reactions to appropriating specific slurs, as the history of the n-word, of "queer" or of "slut" attests. Thus, opposition to the appropriation of the n-word has been present in the African-American cultural space from the beginning (see, among many others, (Asim, 2007) and (Rahman, 2012) for discussion); Brontsema (2004) details the history of the appropriation of "queer", contextualizing the inner struggles within the target community in relation to what is nowadays considered the reclaimed slur par excellence; "slut", on the other hand, is considered an example of a failed attempt at reclaiming a slur, facing strong opposition despite gathering wide support through the famous "slut walks" (see, e.g., Kleinman *et al.*, 2009 or Black Women's Blueprint, 2016).[8] In all these cases, using the slur in the process of appropriation was considered by its opponents as producing harm by reinforcing the dominant, oppressive ideology that the appropriation was supposed to fight against.

On a more general level, Herbert (2015) argues that reclamation is intrinsically a "precarious project". Operating within a speech act framework, Herbert investigates the pragmatic structure of reclamation projects and argues that they face hazards that are distinct from other forms of protest. She points out that reclamation can fail, which in turn leads to harm by reinforcing the very mechanisms of oppression against which the process has started in the first place. Thus, she remarks: "This is the precarious structure of reclamation projects: when successful, reclamation is the subversion of powerful mechanisms of oppression, but when unsuccessful, the act has the ironic force of constituting mechanisms of oppression." (Herbert, 2015, p. 132). However, even though she is wary of reclamation projects, Herbert doesn't argue for banning appropriated or reclaimed uses: "Reclamation is intrinsically hazardous. Yet it can also be worthwhile: even when reclamation projects fail as reclamation, they still may accomplish good. (...) While it will always be precarious, it may on occasion be well worth the risks" (2015, p. 137).[9]

It is interesting to note that appropriated uses of slurs have been seen favourably even by proponents of views that are considered more radical when it comes to cautioning against non-derogatory uses of slurs. For example, as we have seen, under Anderson & Lepore's prohibitionist view, slurs —even in embeddings or quoted— should be banned. However, when it comes to appropriation, they write that "in cases of appropriation, a target group member can opt to use a slur without violating its prohibition because his membership provides a defeasible escape clause" (Anderson & Lepore, 2003, 42). Another class of (radical) views —such as Rappaport's (2020) or Stojnić & Lepore's (2022)— has in common the idea that producing harm is tied to the phonological realization of slurs rather than having to do with their content, either semantic or pragmatic. Yet, when addressing appropriation, Rappaport writes that "in cases of appropriation, it's arguable that a slur can retain its toxicity, despite shifting its conventional use, and such cases may certainly be justifiable" (2020, p. 199), while Stojnić & Lepore (2022) claim that "changes in association over time could also lead to the articulation becoming tied to fully neutral or positive associations, leading to an effective lifting of the ban" (p. 753).

1.2.5. Neutral uses

A different group of non-derogatory uses of slurs consists in them being used neutrally. Anderson (2018), relying on work by Smitherman (2006), points out that within certain communities of African Americans in the U.S.A. the n-word is used to mean "friend", "buddy", "man"; he dubs this type of use *referential*. In Zeman (2022), I have proposed

---

[8] See, however, (Gaucher *et al.*, 2015) for an experimental study showing that the appropriation of "slut" can have positive effects, such as lowering the likelihood of women endorsing rape myths and empowerment leading to positive impact in collective action.

[9] This stance is, of course, consistent with the experimentally proven fact that, when successful, appropriation of a slur has positive effects on the member of the target group. See, among others, (Galinsky *et al.*, 2013).

to interpret a certain use of the ethnic slur "țigan" from Romanian and other Eastern European languages (roughly translated in English as "gipsy") by members of the Roma community as a neutral use; since the aim of such a use is for the members of the community to identify as belonging to a certain ethnic group, I dubbed this use *identificatory*. The main difference between neutral and appropriated uses is that, while the latter aim to put a positive spin on the negative meaning of a slur, the former are used without such an aim.

Neutral uses have not been focused on so far in the literature, so it is difficult to discern what the prevalent attitude towards them would be. While one could extrapolate from what one author or another thinks about non-derogatory uses of slurs in general or about certain types of such uses (e.g., appropriated uses), the results of doing so don't seem promising. However, I take the failure to address such uses as a drawback, since —as I will show— they raise important issues vis-à-vis our stance on banning non-derogatory uses of slurs, both for the authors surveyed so far and for the two who constitute the main focus of this paper.[10]

1.3. A preliminary conclusion

Before moving on to the main argument I want to tackle in this paper, let's take stock. The preceding subsection shows that the harm produced by slurs extends well beyond their derogatory uses. The picture that emerges is that many of the uses of slurs that have been thought to be non-derogatory are not considered as such anymore, while the evidence that they, in fact, produce harm, could form the basis of an argument for banning them.

However, despite this blanket agreement, things get complicated when one considers particular non-derogatory uses and the surveyed authors' attitudes towards them —as even a cursory look at the showcased attitudes and recommendations shows. What seems to transpire is that the authors cited above have slightly different attitudes towards different uses that have been considered non-derogatory. For example, while Anderson & Lepore, Rappaport and Stojnić & Lepore advise against using slurs in quotation, Herbert only urges caution in using them in this way (at the same time offering comprehensive advice on how to mitigate their negative effects; see Herbert, 2017). Appropriated uses have generally seen favourably, even though some authors are keen on pointing out the risks associated with this endeavour (e.g., Herbert, 2015). Thus, even though the authors mentioned converge in the general attitude of cautioning against unrestricted proffering of non-derogatory uses of slurs, there is still some disagreement about the exact nature of the attitude recommended in relation to *particular* types of such uses and under which conditions they should be avoided. This situation seems to preclude both drawing a unitary conclusion vis-à-vis non-derogatory uses of slurs from the ensemble of work investigated and claiming that a general ban on such uses is mandated.

There are other issues to consider as well. For example, if a less lenient attitude is recommended towards one type of non-derogatory use of slurs (say, slurs used in quotation or in reports), but a more permissive one when it comes to another type (say, appropriated uses), then one question that arises is what happens when these two types of uses are combined —for example, when quoting and reporting slurs are part of a process of appropriation. To put it differently, the issue here is that it is unclear what is the "resulting" attitude one should have in such (and similar) cases, and without settling this it is not clear whether the type of use(s) in question should be banned or not. One can claim that, in *this* particular combination of types of uses, the fact that slurs are used, in whatever ways, within the larger umbrella of a process of appropriation makes them free for use. But this is, first, a substantial claim (that appropriation trumps any

[10] Other non-derogatory uses of slurs found in literature are what Hom has called "non-derogatory non- appropriated" uses, illustrated by examples like "Institutions that treat Chinese as chinks are morally depraved." (2008, p. 423), aimed at protesting against or correcting institutional wrongs. There are also metaphorical non-derogatory uses of slurs, as the title of a famous song by John Lennon and Yoko Ono illustrates; the aim of using the n-word here has been not to derogate a racial group (African Americans), but to draw attention to the similar condition of another (women). Jorgensen Bolinger (2020) discusses uses of slurs by competent speakers of a language that are not convinced that the words they use are derogatory. Although interesting, I leave such uses aside, as they have not been discussed in the literature in connection to the issue I'm concerned with here.

other type of use) and, second, something that none of the authors mentioned has explicitly stated. Whatever the answer to particular combinations of types of uses might be, it is not trivial.

Additionally, many of the authors mentioned above have not dealt with neutral uses of slurs. As I stressed above, the difference between neutral and appropriated uses is that in the latter, but not in the former, the resemantization of a slur presupposes a change from negative to positive, which doesn't need to happen in neutral uses. Treating these two types of uses in a similar way might be prima facie reasonable, but (as I will claim in the next section) neutral uses might have certain characteristics that appropriated uses do not have, and thus giving them a common treatment might lead to an incomplete or otherwise inadequate theory. In any case, as with the combination of different types of non-derogatory uses of slurs, this issue has largely remained unaddressed in the literature, and the solution might not be a straightforward one.

Now, many of the authors I engaged with above have, at various times, motivated apermissive attitude towards a type of non-derogatory use or another by claiming that such uses are "worthwhile" and "worth the risk" (Herbert), "justified" or "justifiable" (Rappaport), or governed by a "defeasible escape clause" (Anderson & Lepore). The leading idea behind a permissive attitude thus seems to be that there is something to be gained from a particular use, even if some harm is produced; in most cases, the envisaged advantage is practical or social (in a slogan: a use of a slur is kosher if it is in the service of a good cause). This is certainly a reasonable approach. However, an encompassing theory of what this amounts to, both in general and with respect to particular types of non-derogatory uses, is not easy to find in the literature. What seems to be needed, thus, is a *general* argument of why and when non- derogatory uses of slurs are permissible. In the next section, I turn to a recent theory that aims to offer such an argument, based on the idea of the benefits of using slurs without ignoring their harmful potential, and one that considers their neutral uses as well.

2. The Potential Normalization Argument

In recent work, Alba Moreno Zurita and Eduardo Pérez-Navarro (henceforth "MZ&PN") have urged for caution in relation to any way of using slurs. Their main claim is that "slurs are always potentially harmful, even if some of their occurrences are non-derogatory" (Moreno & Pérez-Navarro, 2021, p. 708). They take this observation to support the stance of not using slurs, even in non-derogatory contexts. Interestingly, and highly relevant for the issue I consider, their position is not a *silentist* one: they do not intend to ban slurs tout court. Instead, what they suggest is that every use of a slur comes with a moral cost, and that it is the speakers' decision (and, hence, responsibility) to determine whether the benefits of using a slur outweigh the cost. What their work adds to the discussion is, first, that they consider non-derogatory uses of slurs that have not been considered in the literature (i.e., *neutral* uses of slurs), and that their argument for being cautious in using slurs non-derogatorily is different from the extant ones in the literature in that it is more general. In this section, I expand on their claims and put forward what I take their main argument to be, which (for reasons that will become clear shortly) I dub the "Potential Normalization Argument".

Thus, MZ&PN start with distinguishing two types of contexts in which slurs can be used: *controlled* contexts and *uncontrolled* contexts. The latter are "contexts in which we do not have enough knowledge of our audience to predict what the uptake of the utterance will be" (Moreno & Pérez-Navarro, 2021, p. 710), due to the fact that most factors that impact the effect of an utterance are beyond our control; the former are contexts in which "we can predict with reasonable accuracy what the consequences of a given utterance will be" (Moreno & Pérez-Navarro, 2021, p. 715), due to our familiarity with the audience. They think that this distinction overlaps with that between derogatory and non-derogatory uses of slurs; that is, while slurs always[11] derogate when used in uncontrolled con-

[11] MZ&PN address cases of uncontrolled contexts in which all the members of the audience take the use of a slur to be non-derogatory by appealing to the idea (which they borrow from Lasersohn, 2007) that "hearers have *every reason* to attribute to [the speaker] a negative attitude toward the target group" (Moreno & Pérez-Navarro, 2021, p. 718). I skip the discussion of what exactly that amounts to, as it doesn't bear on the issues I want to raise.

texts, they can be used non-derogatorily in controlled ones (the claim is not that they always are used like this in such contexts —the use of a slur in a controlled context formed only by members of a hate group is still derogatory). As examples of controlled contexts, MZ&PN provide two scenarios, one involving a pedagogic use of a slur, the other involving an ironic use. Here is how they describe the two situations, as well as their comments on them (labels mine):

> THE PEDAGOGIC SCENARIO
>
> Our son Dani comes home from school and says his friend Y says his other friend X is an S. Later on, we tell Dani he should never say that word again. "What word?", he says. He has not forgotten it, but honestly cannot recall which one of the words he has pronounced we are forbidding him from saying. We feel forced to pronounce "S" in order to make sure he knows what term we are referring to, so we do —we say "We don't call people 'S', that's an ugly thing to say.". We have uttered a slur, even if we have only mentioned it. But we had no other option, and we can be sure that by doing this we have not insulted anyone —if anything, we have prevented Dani from insulting anyone, even if from unintentionally doing so. We know enough about our own son to guarantee that he has understood that we were not insulting anyone. Here, the occurrence of "S" is nonderogatory (Moreno & Pérez-Navarro, 2021, pp. 715-716).

> THE IRONIC SCENARIO
>
> We are a progressive group of friends who would never as much as mention a slur in front of strangers, much less use it to insult a person on grounds of her belonging to a given group. However, we find fun in imitating bigots' mannerisms, and enjoy inner jokes that include ironic uses of "S". We are completely sure that all our friends in the group share our sensibility, and that none of them will take us to aim at insulting anyone. We think it is intuitive to take occurrences of slurs such as these to be nonderogatory, whatever the form of the sentences in which they appear (Moreno & Pérez-Navarro, 2021, p. 716).

The main difference between these two scenarios, MZ&PN claim, is that in THE PEDAGOGIC SCENARIO the slur is mentioned, while in THE IRONIC SCNENARIO, it is used. But, as I mentioned from the outset, and as they also argue, this difference is not relevant when it comes to derogation and thus to producing harm, given that with slurs their derogatory character tends to "scope out" —i.e., hyper-project— both from linguistic and non-linguistic environments.

Now, the next step in their argument is to show that, even with non-derogatory uses of slurs (such as those in the two scenarios presented), there is the potential danger to normalize the use of the slurs in uncontrolled contexts. Thus, the authors show that in relation to THE PEDAGOGIC SCENARIO, by uttering the slur the parents have, inadvertently, given Dani a means to insult the target group —should Dani be so inclined later. Given that the main (or, at least a substantive) harm produced by slurs consists in normalizing a certain way of thinking and speaking about the target group (that is, by normalizing derogation), since the parents have no way of excluding such future uses, even their non-derogatory use of the slur has this normalization potential. The same holds in relation to THE IRONIC SCENARIO: it is entirely possible that, by repeated use, the slur loses its forbidden character and dulls the friends' sensitivity to it, thus making it more likely to be used in uncontrolled contexts. So, to summarize this step in their argument, "no matter how carefully we arrange the current context to make sure that the utterance of a slur does not have the kind of effect we want to avoid, it will facilitate ulterior occurrences of the term. In particular, it will make the slur more likely to appear in uncontrolled contexts in which the utterance of the slur is derogatory" (Moreno & Pérez-Navarro, 2021, p. 716).

Putting all this together, here is what I take their argument to be:

> ***Premise 1***. Normalizing derogation leads to harm.
>
> ***Premise 2***. Uses of slurs in uncontrolled contexts have a high potential to normalize derogation.
>
> ***Partial conclusion 1***: Uses of slurs in uncontrolled context have a high potential to lead to harm.
>
> ***Premise 3***: The potential to normalize derogation is also present with uses of slurs in controlled contexts.
>
> ***Partial conclusion 2***: Uses of slurs in controlled context have a potential to lead to harm.
>
> ***Conclusion***: "[A]ll [uses] of slurs are potentially harmful" (Moreno & Pérez-Navarro 2021, p. 719).

As a final ingredient of their view, and as I already noted, it is important to be clear that MZ&PN are not silentists, as Anderson & Lepore (2013) and others were in the case of, say, slurs used in quotation or embeddings. In other words, while the conclusion of their argument is that all uses of slurs are potentially harmful, this doesn't necessarily support *banning* them across the board. What they hold instead is that each use of a slur comes with a moral cost (as it should be clear by now, *both* in the case of derogatory and non-derogatory uses), and that in some cases "[t]he moral benefits of performing a certain utterance of a slur might outweigh the pervasive moral cost we have described, and so it might be worth it to utter the slur" (Moreno & Pérez-Navarro, 2021, p. 721). Such cases might include the use of the slur in THE PEDAGOGIC SCENARIO, but also uses of slurs in appropriation, in academic contexts,[12] and possibly in the cases pertaining to law and journalism mentioned at the beginning of the paper. As for who bears the responsibility of deciding to use a slur (by carefully weighing the costs and the benefits of using it), MZ&PN are clear that it is the speaker: as they write summarizing their view, "[o]ur point is that uttering a slur always comes at a moral cost, and it is the responsibility of the speaker (...) to assess such cost and decide whether [the slur] is worth [mentioning]." (Moreno & Pérez-Navarro 2021, p. 723).

3. An objection

The Potential Normalization Argument seems to be in good standing, and is consistent with many of the appeals to caution in using slurs non-derogatorily we saw in section 1. Although they rely on two particular scenarios in which slurs are used non-derogatorily, the apport of their argument is more general, and I take MZ&PN to show the limits of a lenient approach to non-derogatory uses of slurs. And even though they don't issue an explicit ban, the point of being cautious, and of thinking of the consequences of such uses, is potently made.

Several objections can be raised to this proposal. To start with, MZ&PN are not entirely clear on the notion of "moral benefits" they are using. Depending on how one understands this notion, a certain consequence of using a slur might count as a benefit or not. Further, since they claim that "performing a certain utterance of a slur might outweigh the pervasive moral cost", this implies that a cost-benefit analysis is performed when a slur is uttered; however, how exactly to perform such an analysis is far from trivial. One could, for example, point out that not all the consequences of a certain act (here, the proffering of a slur) can be known, and so it might be impossible to know whether the benefits outweigh the costs or vice versa. A more general objection is that the view presupposes (or at least comes very close to) a consequentialist outlook: not only has this view been criticised extensively, but there are also many other ethical theories on the market which might apply to the issue of using slurs and possibly yield different results than the consequentialist route the two authors embark on. Finally, a third, more focused objection is that, despite what they assume at the outset, what they show with the two scenarios is that, in fact, there are *no* controlled contexts, leading to the conclusion that the scaffolding they use to support their argument is shaky.

---

[12] The two authors engage in a larger discussion about the effects of slurs in academia, reacting for example to Kukla's (2018) and Herbert's (2017) stances on the matter. As academic contexts are not my focus here, I will not enter that discussion.

However, in what follows I will not push my argument along any of these directions; instead, what I want to focus on is an issue having to do with neutral uses of slurs. The complaint is that, although they consider such uses —which is a step forward in the literature— their reaction to them strikes me as not entirely satisfactory. To be more precise, what I see as problematic in this respect is not the argument *per se*, but one of the accompanying claims they make —namely, that the responsibility of using a slur resides with the speaker.

I start my case by noting that, for some non-derogatory uses to make sense, normalizing derogation is a necessary condition. For example, at least in some cases of appropriated uses, the starting point is to fight back against discrimination, the latter being aided (at least in part) by normalizing derogation of the target group by means of the slur to be appropriated, while the end goal of the process is to normalize the use of the slur without keeping its derogatory meaning. Thus, basically, in appropriation one type of normalization is replaced by another one (the latter devoid of at least *some*, if not all, derogation); this seems to be the chief way of achieving the change in meaning that is required by the goals pursued by the group doing the appropriation. While the range of the group within which such normalization is sought varies (in the case of the n-word, for example, it is the target group, whereas in the case of "queer", the group is everyone), the main goal remains the (arguably gradual) change in meaning. But, quite trivially, for this to happen, normalized derogation by means of the slur has to be in place, with further, repeated, appropriated uses leading to the desired change in meaning. These considerations show two things: that normalization, in itself, is not the (essentially) problematic part in using slurs and that in certain cases normalized derogation is, in fact, a precondition for using slurs in certain ways. But, if this is the case, MZ&PN's argument —which takes the potential normalization of derogation as the main problem with using slurs— loses some of its bite.

A related issue is this. It has been observed by various authors that, in the early stages of appropriation, it is quite difficult to say whether a particular use of a slur is entirely devoid of derogation or still carries it.[13] Additionally, as Brontsema (2004) shows, appropriation is a complex phenomenon than can take many forms, and one dimension of variation is precisely whether the derogatory character of a slur is preserved in appropriation or not (see her comparison between "queer" and the n-word). If these two situations hold, at least for some slurs and in some contexts, then it follows that (quasi-)derogatory uses of slurs are inescapable even in resemantization processes such as appropriation.

Now, MZ&PN's reply is that in appropriation (including the early stages of the process and the types of cases Brontsema focuses on), the benefits of using the slur overweigh the costs, and thus that its use is justified. As we saw above, it is not entirely straightforward how to establish such a claim, given that there is no clear notion of "benefit" the two authors work with. However, even if there were such a notion, the point I want to make is that for other non- derogatory uses of slurs appealing to a "cost-benefit analysis" seems to misrepresent the facts of their very use. For example, a member of a target group using a slur in a neutral way (for example, either using it referentially (in the sense of Anderson, 2018) or in an identificatory manner (as pointed out by Zeman, 2022) seems to happen organically, without the speaker going through a cost-benefit analysis of the sort MZ&PN employ. In the case of the latter, the use of the slur is rather a historical accident, since the target community either didn't have an alternative term or it didn't catch up within the community, and thus it was (or in some cases still is) the only means for some members of the community to identify as part of an ethnic group. This, in turn, casts doubt on the two authors' claim that the responsibility of using a slur resides with the speaker: if using a slur neutrally happens organically without involving a cost- benefit analysis, then it seems exaggerated to ask of speakers from the target communities to pursue one —on pain of being accused of producing potential harm by normalization. Importantly, such uses are crucially different from a derogatory use, which can also happen organically and without involving a cost-benefit analysis (as with people who have been acculturated within a bigoted community), in that it is members of the target group *themselves* that use the slur.

[13] This has led some authors to claim that, at that stage, the slur is ambiguous between a derogatory and a non-derogatory meaning (Jeshion, 2013, 2020; Cepollaro 2020; Jusińska, 2021; Zeman, 2022).

There are two possible replies to this argument.[14] First, the fact that neutral uses happen organically doesn't show that that the speakers from the target group don't go through a cost-benefit analysis. In such cases, the said analysis can be considered to be implicit, or automatic. Second, assuming that speakers using slurs neutrally don't go through a cost-benefit analysis, it doesn't follow that they *shouldn't*: while the former claim is a descriptive/empirical one, the latter is a normative one, and claims of the former sort don't establish claims of the latter sort. In relation to the first reply, one can point out that while it is certainly possible that speakers using slurs neutrally in the spontaneous way mentioned above do, in fact, go through an implicit or automatic cost-benefit analysis, this is something that has to be established by empirical studies in order to be taken seriously. For the likely upshot of speakers going through the analysis in question, even if performed implicitly and automatically, is a significant cognitive burden —or at least one that is greater than that of speakers *not* going through it. As far as I know, no such evidence is available in the case of uses of slurs. Absent such evidence, postulating that the cost-benefit analysis is undertaken by speakers is the more demanding, less economical theory than the one postulating that it is not. So, while I concede the possibility of an implicit and automatic cost-benefit analysis done by speakers in the relevant cases, since it is the proponents of this reply that make this strong claim, it befalls on them to back it up.

In relation to the second reply, while I agree that the descriptive/empirical claim doesn't imply the normative one, it merits stressing that they are connected to each other in a significant way. Suppose one wanted to argue that speakers of neutral uses of slurs should perform a cost-benefit analysis, despite the fact that they do not appear to do so. Certainly, for this claim to be true, these speakers should be capable of performing, or be in a position to perform, such an analysis. There is no doubt that generally these speakers have that ability, given that what it amounts to is merely being reflective and thinking of the consequences of one's words on the well-being of others. However, we have to be careful here. I think that, even if one accepts that the claim that the responsibility of going through a cost-benefit analysis lies with the speaker is a normative one, the case of neutral uses of slurs that happen spontaneously opens up some different, and pressing, *normative* questions. One type of question that appears is about *who* judges that speakers belonging to a certain group should make the cost-benefit analysis and are thus justified or not in using the slur. The social position of those who ask such a thing from speakers (including theorists!) is relevant: if, for example, the speaker is a member of a disadvantaged group, being asked to perform the cost-benefit analysis by a member of an advantaged group raises the threat of *paternalism*: the reality of the life of a member of a disadvantaged group (which might be poverty, lack of education, lack of opportunities, etc.) could simply be erased in the service of casting a general moral judgment. Things might be different when the speaker is part of an advantaged group, for which perhaps the moral stringencies are backed up by access to better living conditions, education, opportunities, etc.

Now, the threat of paternalism doesn't establish that speakers from a disadvantaged group shouldn't be asked to go through the cost-benefit analysis envisaged by MZ&PN, for one could make a distinction between contextual or circumstantial factors that annul a moral obligation and ones that merely make the subject warranted (or blameless) in not fulfilling it. One might thus argue that no circumstantial factors annul a moral obligation or that those mentioned above (poverty, lack of education, lack of opportunities, etc.) are not of the required kind. This touches on delicate ethical issues, and the current paper is not the place to tackle them. The point that I contend myself of making is that whether the threat of paternalism gets in the way of considering asking speakers from disadvantaged groups to go through a cost- benefit analysis before using a slur depends on taking a stance on the ethical issues alluded to. I maintain that this is a substantial burden, one that I'm not sure MZ&PN are ready to take on. In any case, the idea of the social position of those insisting on the cost-benefit analysis seems to me to bring back the issue of the reasonableness of asking speakers from disadvantaged groups to pay close attention to their use of slurs —something that MZ&PN don't fully consider.[15]

---

[14] Both suggested by reviewers, whom I thank.

[15] To better illustrate the type of case I have in mind, let me briefly describe the situation of Roma ethnics in Romania. As many as 650.000 (official number), Roma ethnics have been slaves for several centuries and even after they have been freed, they have continued to live in

A different, less contentious answer to the normative question of who judges that speakers belonging to a certain group should make the cost-benefit analysis and are thus justified or not in using a slur —one that is more sensitive to the social position issue— is this: *other members of that group*. One may argue that other members of a disadvantaged group might be the ones who are better positioned to ask a fellow member to be careful with their use of a slur. Adopting this idea would also explain various intra-group disagreements about whether and how slurs targeting the members of a group should be used by those members *themselves* —as the history of many slurs attests.[16] The difference between this case and the previous one is that the decision of whether and how to use a slur belongs to the target community itself, which should have a say on how language that targets them should be used. Again, as before, and without going into more details, considerations along such lines are not part of MZ&PN's argument itself —but they do seem to affect its plausibility. I think it's fair to say that there is at least a *tension* between the conclusion of the argument, attributing responsibility to the speakers for using a slur, and the neutral uses of slurs focused on.

Now, the two authors address some of these concerns in their papers, having this to say:

> Other nonderogatory uses of slurs that have recently been described are referential (Anderson, 2018) and identificatory (Zeman, 2022). We take these uses to take place in controlled contexts too, as the speaker's group membership is salient enough for her to be confident that the audience will understand that she did not mean to insult, just like happened in Dani's case (Moreno & Pérez-Navarro, 2021, p. 716).

> Referential and identificatory uses of slurs are closer to pedagogic mentions than to ironic uses in this respect. Like with pedagogic mentions, however, there is still the risk that these uses facilitate ulterior occurrences of the term in contexts in which the group membership of the speaker, although salient, does not make the audience understand such occurrences as nonderogatory because the speaker does not belong to the target group (Moreno & Pérez-Navarro, 2021, p. 720).

While these quotes might establish that referential and identificatory uses of slurs have a potential to normalize derogation, they don't address the tension I mentioned above. For even if one considers referential and identificatory uses to happen in controlled contexts, and even if one agrees that they have the potential of being used in uncontrolled ones, there is still no clear answer to the responsibility issue. So, while the conclusion of the Potential Normalization Argument ("all uses of slurs are potentially harmful") remains in place, since the attribution of responsibility goes through a cost-benefit analysis whose reasonability is not unquestioned in cases of neutral uses of slurs, not much follows from it in practical terms. In other words, MZ&PN's analysis doesn't provide an answer to the (normative) question "should *neutral* uses of slurs be banned?", which is essential to answer the larger question of whether all non- derogatory uses of slurs should be banned.

Moreno Zurita (2023) independently considers the two types of neutral uses I marshalled against MZ&PN's view and anticipates some of the issues I raise, as witnessed by the following paragraph:

> Although I have no developed position on this, and only a mere intuition, I do not feel entirely comfortable talking about the harmful potential of appropriated, referential and identificatory uses. I believe that, in these cases, the normalising potential of slurs is irrelevant and that the moral cost assumed, if any for these cases, will always be justified. I just don't think I have

---

poverty. While they are in general better positioned socially and economically in present times, many members of the community still live in suboptimal conditions and are systematically discriminated. The slur used to derogate them, "țigan", has been introduced as an exonym by the ruling classes and is still widely used. Interestingly, however, as discussed in Zeman (2022), a significant number of Roma ethnics use the slur as a way to identify themselves as belonging to the group; such a use is, arguably, non-derogatory. One reason for resorting to the slur for purposes of identification is that some of the Roma ethnics don't speak Romani, their original language, and so have (and had) little choice about the term. Unsurprisingly, I find it unreasonable (and paternalistic) for non-Roma to insist that such members go through a cost-benefit analysis every time they use (non-derogatorily) the term.

[16] See the quick discussion in section 1, subsection 1.2.4.

> the right to say, even if I belonged to one, whether the derogated groups commit any kind of infringement, however minor and justified (Moreno Zurita, 2023, p. 143, footnote 4).

While this is no doubt a morally admirable stance, again, it doesn't get to the heart of the matter. For, on the one hand, the issue is whether there is a cost-benefit analysis present in the case of referential and identificatory uses in the first place, not whether the cost is assumed or the risk of normalization justified. On the other hand, if we take the issue to be a normative one, Moreno Zurita's conclusion seems to be too strong: the problem I pointed out is not that the moral cost is always justified in appropriated, referential and identificatory uses of slurs, but that it depends, in ways unexplored by MZ&PN, on *who* is asking the speakers to police their use. Additionally, saying that in these cases "the normalizing potential is irrelevant" amounts in my view to a serious concession, limiting the main thesis to only a subclass of uses of slurs. This is unsatisfactory because, while derogatory uses are those that best fulfil slurs' main function and are perhaps the most widespread, many uses of slurs are non-derogatory.

4. Summary and conclusions

In this paper, I have engaged with Alba Moreno Zurita and Eduardo Pérez-Navarro's view on the harms of non-derogatory uses of slurs. The two authors urge caution in relation to *any* use of slurs based on their potential —both in what the authors distinguish as controlled and uncontrolled contexts— to normalize derogation. Before getting to their view, I have set the scene by presenting the main types of uses that have been taken to be non-derogatory and the more recent reactions to them in the literature (section 1). The picture that emerged was that non-derogatory uses of slurs can have pernicious effects on the target group, on a par with derogatory ones. Yet, I claimed, due to the variation in the exact attitude towards such uses various authors have recommended, no consensus —ether about a specific type of use or about non-derogatory uses in general— can be extracted, except the rather vague claim that using slurs is permissible under the condition that they lead to some practical or social gain. I turned then to MZ&PN's view, which I take to have the level of generality required.

Against this background, I presented the two authors' main argument —the "Potential Normalization Argument", as I called it (section 2). While I deemed the argument to be in good standing, I have offered an objection to their treatment of neutral uses of slurs (a subspecies of non-derogatory uses) targeting an additional claim they make —namely, that the responsibility for using slurs lies with the speakers (section 3). I have shown that in using slurs neutrally, speakers don't seem to employ the cost-benefit analysis MZ&PN claims should lie at the basis of all uses of slurs, which leads to the claim that taking such speakers to be responsible is not reasonable. I have then considered two replies to this objection, which in turn led to considerations about the role of empirical studies about the use of slurs and about who exactly has the moral standing to demand speakers to make the cost-benefit analysis. However, whether my objection to MZ&PN leads to the rejection of their argument or merely prompts further fine-tuning remains to be seen. While I myself don't offer any consideration for or against banning non-derogatory uses of slurs, I take my contribution to be that of bringing into focus certain (important) aspects not covered by their argument. And while the two authors have moved the issue forward by offering a more general argument than those found in literature and by considering neutral uses of slurs, this (critical) paper is an invitation to think things even further.

Acknowledgments

I thank Alba Moreno Zurita, Eduardo Pérez-Navarro and Sergio Guerra for extensive discussions on the issues tackled in the paper. I also thank the audiences at the *Hate Speech, Fake News and Freedom of Speech* workshop, 18-19.07.2023,

University of Rijeka, the *New Directions in Social Philosophy of Language* workshop, 13-14.11.2024, NOVA University of Lisbon and the *Hybrid Representations* reading group, 10.01.2025, University of Barcelona, where previous versions of this paper were presented. I acknowledge the financial help of a Principal Investigator grant (2023.05952.CEECIND) from Fundação para a Ciência e a Tecnologia - FCT (Portugal).

REFERENCES

Anderson, L. (2015). Racist humor. *Philosophy Compass, 10*(8), 501-509. https://doi.org/10.1111/phc3.12240

Anderson, L. (2018). Calling, addressing, and appropriation. In D. Sosa (Ed.), *Bad words: Philosophical perspectives on slurs* (pp. 6-28). Oxford University Press.

Anderson, L., & Lepore, E. (2013). Slurring words. *Noûs, 47*, 25-48. https://doi.org/10.1111/j.1468-0068.2010.00820.x

Asim, J. (2007). *The N word: Who can say it, who shouldn't, and why*. Houghton Mifflin Harcourt.

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics, 66*, 35-44. https://doi.org/10.1016/j.pragma.2014.02.009

Black Women's Blueprint. (2016). An open letter from Black women to the SlutWalk. *Gender and Society, 30*(1), 9-14. https://doi.org/10.1177/0891243215611868

Blackburn, S. (1984). *Spreading the word*. Oxford University Press.

Brontsema, R. (2004). A queer revolution: Reconceptualizing the debate over linguistic reclamation. *Colorado Research in Linguistics, 17*(1), 1-17.

Carnaghi, A., & Maass, A. (2008). Derogatory language in intergroup context: Are "gay" and "fag" synonymous? In Y. Kashima, K. Fiedler, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to the formation, maintenance, and transformation of stereotypes* (pp. 117-134). Lawrence Erlbaum Associates.

Cepollaro, B. (2020). *Slurs and thick terms: When language encodes values*. Lexington Books.

Cepollaro, B., Sulpizio, S., & Bianchi, C. (2019). How bad is it to report a slur? An empirical investigation. *Journal of Pragmatics, 146*, 32-42. https://doi.org/10.1016/j.pragma.2019.03.012

Cepollaro, B., Sulpizio, S., Bianchi, C., & Stojanovic, I. (2024). Slurs in quarantine. *Mind & Language, 39*(3), 381–396. https://doi.org/10.1111/mila.12492

Cepollaro, B., & Zeman, D. (2020). Editors' introduction: The challenge from non-derogatory uses of slurs. *Grazer Philosophische Studien, 97*(1), 1-10.

Croom, A. (2011). Slurs. *Language Sciences, 33*(3), 343-358. https://doi.org/10.1016/j.langsci.2010.11.005

Davis, C., & McCready, E. (2020). The instability of slurs. *Grazer Philosophische Studien, 97*, 63-85. https://doi.org/10.1163/18756735-09701005

Dummett, M. (2007). Reply to Picardi. In B. McGuinness & G. Oliveri (Eds.), *The philosophy of Michael Dummett* (pp. 521-529). Open Court.

Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2016). Not "just words": Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology, 46*(2), 237-248. https://doi.org/10.1002/ejsp.2148

Galinsky, A. D., Wang, C. S., Whitson, J. A., Anicich, E. M., Hugenberg, K., & Bodenhausen, G. V. (2013). The reappropriation of stigmatizing labels: The reciprocal relationship between power and self-labeling. *Psychological Science, 20*(10), 2020-2029. https://doi.org/10.1177/0956797613482943

García-Carpintero, M. (2004). The deferred ostension theory of quotation. *Noûs, 38*(4), 674-692. https://doi.org/10.1111/j.0029-4624.2004.00488.x

Gaucher, D., Hunt, B., & Sinclair, L. (2015). Can pejorative terms ever lead to positive social consequences? The case of SlutWalk. *Language Sciences, 52*, 121-130. https://doi.org/10.1016/j.langsci.2015.03.005

Herbert, C. (2015). Precarious projects: The performative structure of reclamation. *Language Sciences, 52*, 131-138. https://doi.org/10.1016/j.langsci.2015.05.002

Herbert, C. (2017). *Exclusionary speech and constructions of community* (Doctoral dissertation). Georgetown University.

Hom, C. (2008). The semantics of racial epithets. *The Journal of Philosophy, 105*(8), 416-440. https://doi.org/10.5840/jphil2008105834

Hornsby, J. (2001). Meaning and uselessness: How to think about derogatory words. *Midwest Studies in Philosophy, 25*(1), 128-141. https://doi.org/10.1111/1475-4975.00042

Jeshion, R. (2013). Expressivism and the offensiveness of slurs. *Philosophical Perspectives, 27*, 231-259. https://doi.org/10.1111/phpe.12027

Jeshion, R. (2020). Pride and prejudiced. *Grazer Philosophische Studien, 97*(1), 106-137. https://doi.org/10.1163/18756735-09701007

Jorgensen Bolinger, R. (2020). Contested slurs: Delimiting the linguistic community. *Grazer Philosophische Studien, 97*(1), 11-30. https://doi.org/10.1163/18756735-09701003

Jusińska, Z. (2021). Slur reclamation – Polysemy, echo, or both? *Organon F, 28*(3), 689-707. https://doi.org/10.31577/orgf.2021.28310

Kennedy, R. (2003). *Nigger: The strange career of a troublesome word*. Penguin Random House.

Kleinman, S., Ezzell, M. B., & Frost, C. A. (2009). Reclaiming critical analysis: The social harms of 'bitch.' *Sociological Analysis, 3*(1), 46-68.

Kukla, R. (2018). Slurs, interpellation, and ideology. *The Southern Journal of Philosophy, 56*(S1), 7-32.

Lasersohn, P. (2007). Expressives, perspective, and presupposition. *Theoretical Linguistics, 33*(2), 223-230. https://doi.org/10.1515/TL.2007.015

Livingstone Smith, D. (2011). *Less than human*. St. Martin's Press.

Livingstone Smith, D. (2020). *On inhumanity: Dehumanization and how to resist it*. Oxford University Press.

Livingstone Smith, D. (2021). *Making monsters: The uncanny power of dehumanization*. Harvard University Press.

Moreno, A., & Pérez-Navarro, E. (2021). Beyond the conversation: The pervasive danger of slurs. *Organon F, 28*(3), 708-725. https://doi.org/10.31577/orgf.2021.28311

Moreno Zurita, A. (2023). *Hiding in plain sight: A non-propositional presuppositional account of slurs* (Doctoral dissertation). University of Granada.

Naylor, G. (1986, March 30). The meanings of a word. *The New York Times*.

Popa-Wyatt, M., & Wyatt, J. L. (2018). Slurs, roles, and power. *Philosophical Studies, 175*, 2879-2906. https://doi.org/10.1007/s11098-017-0986-2

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics, 33*(2), 165-198. https://doi.org/10.1515/TL.2007.011

Rahman, J. (2012). The N word: Its history and use in the African American community. *Journal of English Linguistics, 40*(2), 137-171. https://doi.org/10.1177/0075424211414807

Rappaport, J. (2020). Slurs and toxicity: It's not about meaning. *Grazer Philosophische Studien, 97*(1), 177-202. https://doi.org/10.1163/18756735-09701010

Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy, 26*(1), 29-120. https://doi.org/10.1023/A:1022225203544

Smitherman, G. (2006). *Word from the mother: Language and African Americans*. Routledge.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379-440.

Stojnić, U., & Lepore, E. (2022). Inescapable articulations: Vessels of lexical effects. *Noûs, 56*, 742-760.

Swanson, E. (forthcoming). Slurs and ideology. In R. Celikates, S. Haslanger, & J. Stanley (Eds.), *Analysing ideology*. Oxford University Press.

Tirrell, L. (2012). Genocidal language games. In I. Maitra & M. K. McGowan (Eds.), *Speech and harm: Controversies over free speech* (pp. 174-221). Oxford University Press.

Williamson, T. (2009). Reference, inference, and the semantics of pejoratives. In J. Almog & P. Leonardi (Eds.), *The philosophy of David Kaplan* (pp. 137-158). Oxford University Press.

Zeman, D. (2022). A rich-lexicon theory of slurs and their uses. *Inquiry, 65*(7), 942-966. https://doi.org/10.1080/0020174X.2021.1903552

Dan Zeman specializes in philosophy of language and formal semantics. He is interested in accounts of various natural language expressions such as indexicals, predicates of taste, aesthetic and moral terms, epistemic modals, slurs, expressives, code words and gender terms. He has published a significant number of papers in journals and collective volumes, and has done editing work as well. He is one of the coordinators of the *Slurring Terms Across Languages* international network.

Address: Institute of Philosophy, Faculty of Letters, University of Porto, Via Panorâmica, s/n, 4150-564, Porto, Portugal. E-mail: zeman@letras.up.pt – ORCID: https://orcid.org/0000-0001-5620-0042

Theoria, 2025, 40(1), 47-67
https://doi.org/10.1387/theoria.26444

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

The internal structure of dual character concepts: A corpus-based study of scientist

(La estructura interna de los conceptos de doble carácter: Un estudio basado en corpus de científico)

David Bordonaba-Plou*
Universidad Complutense de Madrid
https://orcid.org/0000-0002-0788-9733

Keywords

Dual character concepts
Descriptive dimension
Normative dimension
Experimental philosophy
Corpus methods

ABSTRACT: Over the past decade, there has been a growing interest in dual character concepts (DCCs). These concepts are defined by their internal structures, which consist of two distinct dimensions: a descriptive and an independent normative dimension. However, a more in-depth exploration of their internal structures is still needed. This article examines the internal structure of one DCC that has garnered significant attention in the literature, SCIENTIST. First, I analyze the components of the different dimensions of this concept. Second, I explore the interaction between these two dimensions. To do so, I investigate SCIENTIST in the enTenTen20 corpus using Sketch Engine, focusing on the expressions "good scientist" and "true scientist", as the literature suggests they interact more directly with the descriptive and normative dimensions, respectively. The findings from this investigation offer valuable insights for studying other DCCs, as the results suggest, among others, the following key points: first, that the complexity of the two dimensions of SCIENTIST is greater than previously recognized; and second, contrary to what is agreed, both the descriptive and the normative dimension interact with "good" and "true," which implies that both expressions can be used to make the two types of normative evaluation proper of DCCs.

Palabras clave

Conceptos de carácter doble
Dimensión descriptiva
Dimensión normativa
Filosofía experimental
Métodos de corpus

RESUMEN: Durante la última década, ha habido un interés creciente en los conceptos de carácter doble (CCDs). Estos conceptos se definen por sus estructuras internas, que constan de dos dimensiones distintas: una dimensión descriptiva y otra dimensión normativa independiente. Sin embargo, todavía es necesaria una exploración más profunda de sus estructuras internas. Este artículo examina la estructura interna de un CCD que ha suscitado gran atención en la literatura, CIENTÍFICO. En primer lugar, analizo los componentes de las distintas dimensiones de este concepto. En segundo lugar, exploro la interacción entre estas dos dimensiones. Para ello, investigo CIENTÍFICO en el corpus enTenTen20 utilizando Sketch Engine, centrándome en las expresiones "buen científico" y "verdadero científico", ya que la literatura sugiere que interactúan más directamente con las dimensiones descriptiva y normativa, respectivamente. Los resultados de esta investigación ofrecen información valiosa para el estudio de otros CCDs, ya que los resultados sugieren, entre otros, los siguientes puntos clave: en primer lugar, que la complejidad de las dos dimensiones de CIENTÍFICO es mayor de lo que se había reconocido anteriormente; y en segundo lugar, en contra de lo que se piensa, tanto la dimensión descriptiva como la normativa interactúan con "bueno" y "verdadero", lo que implica que ambas expresiones pueden utilizarse para realizar los dos tipos de evaluación normativa propia de los CCDs.

* ***Correspondence to:*** David Bordonaba-Plou. Universidad Complutense de Madrid, Facultad de Filosofía, Dept. de Lógica y Filosofía Teórica, Despacho B-09, Edif. B, Calle del Prof. Aranguren, s/n, Ciudad Universitaria, 28040, Madrid, Spain – davbordo@ucm.es – https://orcid.org/0000-0002-0788-9733

How to cite: Bordonaba-Plou, David (2025). «The internal structure of dual character concepts: A corpus-based study of SCIENTIST»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 47-67. (https://doi.org/10.1387/theoria.26444).

Received: 05 June, 2024; Final version: 24 April, 2025.

ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

This work is licensed under a *Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International* License

1. Introduction

Over the past few decades, there has been a growing interest in those concepts whose meaning combines descriptive and non-descriptive features. Thick concepts (see Williams, 1985; Dancy, 1995; Väyrynen, 2011, 2013; Kirchin, 2013) represent perhaps the most well-known and extensively studied example. However, in the last ten years, dual character concepts (henceforth, DCCs) have also garnered significant attention (see Knobe *et al.*, 2013; Del Pinal & Reuter, 2017; Reuter, 2019; Phillips & Plunkett, 2023). According to Knobe *et al.* (2013, p. 243), DCCs "involve two ways of characterizing their instances, and thus two ways of determining category membership." First, a set of specific features; second, an underlying abstract value. In other words, the internal structure of DCCs is composed of two different dimensions. A paradigmatic example of a DCC is the concept SCIENTIST,[1] since most authors discussing the meaning of DCCs focus on this concept. As explained, SCIENTIST encompasses concrete features such as "Conducting experiments, Analyzing data, Developing theories, Writing papers" (Knobe *et al.*, 2013, p. 243) and also an abstract value like "the pursuit of scientific knowledge" (Knobe *et al.*, 2013, p. 243) or "the quest for impartial truth" (Del Pinal & Reuter, 2017, p. 479).

An important feature of DCCs is that the two dimensions are related but independent (Knobe *et al.*, 2013, 243; Del Pinal & Reuter, 2017, pp. 477, 492; Reuter, 2019, p. 1; Guo *et al.*, 2021, p. 2). This means someone can be considered a scientist according to the descriptive dimension (if s/he possesses the concrete features), but not according to the normative dimension (if s/he does not instantiate the abstract value), and vice versa. For example, take Emma, an astronomer who spends a lot of time gathering data and conducting experiments, but solely to enrich herself. Despite her specific motives, we would still say that Emma is a scientist because her professional activities include many tasks typical of a scientist. In this sense, Emma only fulfills the descriptive dimension. Now consider Peter, who has never received scientific training, but his life is deeply rooted in the pursuit of objective truth. In this case, we might say that, even though Peter does not perform any of the tasks usually performed by scientists, Peter is a scientist because his way of life aligns with the abstract value that defines a scientist. In this sense, Peter only fulfills the normative dimension. Before proceeding, a brief clarification is in order.

DCCs are concepts, and concepts can be defined in different ways. The most widespread position on DCCs is the so-called "lexical-semantic view," (Baumgartner, 2024, p. 727) which argues that DCCs are lexical concepts. One of the most defining features of lexical concepts is that they are elements of mental grammar, i.e., mental representations (Fodor, 1975); in other words, they "are abstractions over multiple instances of language use." (Evans, 2009, p. 128). However, concepts can also be defined as abstract objects or Fregean senses (Frege, 1891/1997, 1892/1997). A final option is to say that concepts are abilities (Wittgenstein, 1953; Kenny, 2010). That is, someone possesses the concept CAT when she has mastered the rules for using the word "cat."

My approach is grounded on ordinary language philosophy (Wittgenstein, 1953; Austin, 1962) for several reasons, the most important of which is that I do not want my thesis or my theories to lose sight of the phenomenon I want to study. As Baz (2012, p. 4) highlights, not paying attention to the real use of the terms we are interested in, "the philosopher risks having his theory lose contact with the world it is supposed to help us illuminate." Consequently, instead of looking for the meaning that allows us to determine the circumstances of application of a word, I examine the usages to determine the meaning of the word. Since this paper focuses on language use —specifically, how competent speakers use expressions like "scientist"— I adopt the third perspective on concepts, concepts as abilities. In this view, a DCC like SCIENTIST is simply the ability to apply the term "scientist" according to the rules of use established within a linguistic community. Mastering these rules likely involves a broader range of uses than those required for purely descriptive concepts (e.g., ROUND), as the former, unlike the latter, encompasses both descriptive and normative uses. In other words, there will be cases where something can be considered a scientist without possessing the descriptive features

[1] I follow the practice of using small caps for concepts.

(or, if you prefer, without fulfilling the descriptive dimension) and cases where something can be considered a scientist without possessing the normative features (or, if you prefer, without fulfilling the normative dimension).

There has been much progress focusing on DCCs since they have been used to shed light on different topics, for example, gender terms and generics (Leslie, 2015; Guo *et al.*, 2021), natural kind terms (Tobia *et al.*, 2020), personal identity (Knobe, 2022), art (Liao *et al.*, 2020), or law (Flanagan & Hannikainen, 2022; Almeida *et al.*, 2023). However, many questions remain about their internal structure, such as which components comprise the descriptive and normative dimensions, and how these two dimensions interrelate. In this regard, Phillips & Plunkett (2023, p. 341) say that "a key debate is about what exactly the "internal structure" of dual-character concepts consists in." In a similar vein, Reuter (2019, p. 4) notes that "[w]hereas many controversies remain in regard to detailing the normative content of dual character concepts, a further issue arises as to if and how strongly both dimensions interact."

This paper explores the internal structure of DCCs, aiming to illuminate the interplay between the descriptive and the normative dimensions. I will use the concept SCIENTIST as a model. First, I will analyze the components of the different dimensions of this concept, assessing how well the elements identified in the literature represent it. Second, I will investigate the interaction between these two dimensions. Rather than employing questionnaires, the most widespread methodology in experimental philosophy and studies on DCCs, I will use corpus methods. Specifically, I will examine the concept SCIENTIST in the enTenTen20 corpus (see Jakubíček *et al.*, 2013) in Sketch Engine, searching for the expressions "good scientist" and "true scientist", as the literature suggests they interact more directly with the descriptive and normative dimension, respectively.

I proceed as follows: Section 2 introduces DCCs, focusing on various proposals for operationalizing the two dimensions. Section 3 details the methods and materials used in the study. Section 4 presents the results of the corpus analysis for the expressions "good scientist" and "true scientist." Finally, Section 5 discusses the significance of the results for the literature on DCCs.

2. Explaining the two dimensions of DCCs

There are different proposals on how to operationalize the two dimensions of DCCs. Knobe *et al.* (2013) identify the descriptive dimension with a set of concrete or specific features and the normative dimension with an underlying abstract value. If you were to ask someone what it means to be a scientist, they would most likely start by giving you a list of concrete features that most scientists tend to exhibit, for example, postulate hypotheses about some phenomenon, collect data to perform experiments to prove or disprove such hypothesis, or write journal articles to disseminate the results. However, as Knobe and his colleagues point out, this list of concrete features would not be arbitrary, but would all be related to the same abstract value, in this case, something like the search for objectivity or unbiased truth. In short, speakers associate with DCCs certain concrete features that together they realize an abstract value.

Leslie (2015) and Del Pinal & Reuter (2017), however, focus on social roles, linking the descriptive dimension to typical functions of a role (e.g., a scientist learns mathematics or conduct experiments) and the normative dimension to idealized functions of a role (e.g., a scientist seeks objective knowledge or impartial truth).[2]

As mentioned in the Introduction, one of the defining features of DCCs is that the two dimensions are related, but independent. This means that while there is a connection between the descriptive and the normative dimension, an indi-

[2] It should be noted that, although both focus on the idea of social roles, there are differences between the two proposals. For Leslie (2015, p. 130), if someone "is a member of a social kind, and that social kind has a particular primary role or function, then there is a prima facie obligation to fulfill that role or function." However, for Del Pinal & Reuter (2017, p. 479), "what matters most is not whether someone actually fulfills the basic function to some non-trivial degree, but rather whether someone is committed to fulfill it."

vidual does not need to exhibit characteristics of both to be classified under a particular DCC.[3] Remember our example of Emma and Peter. Emma exclusively exhibited characteristics of the descriptive dimension, yet we still considered her a scientist. Conversely, Peter exclusively exhibited characteristics of the normative dimension, yet he too was considered a scientist.

Although more evidence is needed on how the two dimensions interact, several proposals in this regard can be found in the literature. Firstly, the two-senses approach states that people display a specific intuition pattern, showing they have two different characterizations for DCCs (for example, SCIENTIST), but not for single character concepts (for example, OPTICIAN). According to the proponents of the two-senses approach, it is obvious that the following conversation is commonplace:

1) Peter: Emma is not a scientist. I know that many people think she is a scientist, but, if you think about what it really means to be a scientist, she's not a scientist.
2) Maria: How can you say that Emma is not a scientist? She works in a lab and spends her days conducting experiments.
3) Peter: Okay, that's not what I meant. What I'm trying to say is that her actions are not driven by a genuine, impartial pursuit of knowledge. The only thing that matters to her is money.

However, according to the proponents of the two-senses approach, similar conversations involving single character concepts do not sound natural. Consider the same conversation but with SCIENTIST replaced by OPTICIAN, a single character concept:

4) Peter: Emma is not an optician. I know that many people think she is an optician, but, if you think about what it really means to be an optician, she's not an optician.
5) Maria: How can you say that Emma is not an optician? She works in an optician's shop and spends her days conducting eye exams.
6) Peter: Okay, that's not what I meant. What I'm trying to say is that she is not motivated to help humanity to see things in a better light. The only thing that matters to her is money.

According to the two senses approach, most people would accept 1)-3), but would not accept 4)-6). This is because SCIENTIST, unlike OPTICIAN, is a DCC. In this sense, and since the two dimensions of DCCs are independent, they "allow for a double-dissociation in categorizations, such that descriptive categorizations may fail to align with normative categorizations" (Baumgartner, 2024, p. 728). That is, although the two dimensions often overlap, it is possible, depending on the features of the object to be categorized, to reach dissociated categorization judgments. Consider the following example:

7) My cousin Peter is 6 years old, and he is already a true scientist; not like Yoshihiro Sato.

---

[3] Some studies show that the two dimensions are not totally independent. For example, Baumgartner (2024) shows that the abstract values of the normative dimension have another function beyond the categorization function mentioned in the literature, "the *injunctive function* of the normative dimension," which determines what should count as a member of the category. Regarding the categorization function, both dimensions are independent; something can be a member of the category not fulfilling the descriptive criteria. Remember Peter in the Introduction's example. He has not met any of the descriptive criteria, but still, he is a scientist. However, regarding the injunctive function, something cannot be a member of the category if it does not fulfill the descriptive criteria. Remember now Emma in the same example and think if it makes sense to say of her that she ought to be a true scientist. Now, would still be ok to say the same of Emma supposing she would not have met the descriptive criteria? The answer is in the negative. This shows that "injunctions based on the normative dimension of dual character concepts are dependent on descriptive class membership." (Baumgartner, 2024, p. 478).

Someone who utters (7) would say that Peter is a scientist because he fulfills the normative dimension, even though he does not fulfill the descriptive dimension. At the same time, they would say that Yoshihiro Sato is not a scientist because, although he fulfilled the descriptive dimension, he did not fulfill the normative one.

Secondly, several authors argue that while the normative dimension interacts with modifiers like "true" or "real," it does not seem to interact with modifiers such as "good." Leslie (2015, p. 117) observes that "it is natural to suppose that the adjectives "true" and "real" tend to select the normative sense." In a similar vein, Del Pinal & Reuter (2017, p. 479) say that "modifiers such as *true* in colloquial expressions such as *John is not a true father*, [...] seem to operate (not necessarily exclusively) on this normative dimension." Liao *et al.* (2020, p. 108) suggest that "[i]n general, one can pick out this second criterion [the abstract values] by adding the modifier 'true'." Guo *et al.* (2021, p. 7) explore "whether gender concepts resemble dual-character concepts in the extent to which they can be described with "true" (a true man), which concerns abstract traits, as well as "good" (a good man), which concerns concrete traits." Flanagan & Hannikainen (2022, p. 169) defend that "most people have been found to agree that a postdoctoral researcher who is employed to run experimental studies but who is completely uninterested in her findings is clearly a scientist in some sense but is not a true scientist at all." Finally, Baumgartner (2024, p. 728) says that "specific modifiers like "true" or "truly" (and to a lesser extent "real/really") can be utilized to highlight the normative dimension of terms expressing dual character concepts." To illustrate, consider the following sentences:

8) Emma is a good scientist.
9) Emma is a true scientist.

Someone who utters 8) will communicate that Emma possesses certain characteristics necessary for proper scientific practice, for example, the analysis of data or the conducting of experiments. However, someone who utters 9) will communicate that Emma's way of life is in line with an abstract value to whose attainment the scientific practice is oriented.

In summary, there are two different normative evaluations associated with DCCs, which can be highlighted by using the expressions "good" and "true," and, from these two expressions, the latter is often used to highlight the normative dimension. However, as discussed above, further studies investigating the specific composition of these two dimensions or how they interact are needed.

This study fills this gap by focusing on a specific DCC, SCIENTIST, empirically investigating their two dimensions through the expressions "good scientist" and "true scientist." A significant finding from the present study is that, contrary to what is widely agreed, both the descriptive and the normative dimension interact with "good" and "true," which implies that both expressions can be used to make the two types of normative evaluation. However, before detailing the results of the analysis, in the next section, I will present the methods and materials used for this purpose.

3. Methods and Materials

The use of linguistic corpora in analytic philosophy has been defended in recent years (see Bluhm, 2013, 2016; Caton, 2020). Besides, analytic philosophers have fruitfully used corpus methods to investigate different philosophical topics, for example, the role of intuitions (Andow, 2015a, 2015b; Ashton & Mizhari, 2018; Bordonaba-Plou, 2021), the philosopher's use of "know" (Hansen *et al.*, 2021), meta-argumentation (Hinton, 2021), causal attributions (Sytsma *et al.*, 2019), mathematical explanation (Mejía-Ramos *et al.*, 2019), the expression of values in obituaries (Alfano *et al.*, 2018), or the insularity of Anglophone philosophy (Schwitzgebel *et al.*, 2018), among others.

As said in the Introduction, this work investigates the internal structure of SCIENTIST by examining how competent speakers employ "scientist." To accomplish this, I will perform a corpus analysis of the expressions "good scientist" and "true scientist" in the enTenTen20 corpus (see Jakubíček *et al.*, 2013) in Sketch Engine. The enTenTen20 is a sam-

ple corpus, meaning it aims "to represent a particular type of language" (McEnery & Hardie, 2012, p. 8); in this case, the English language. Its size is enormous, 36 billion words, featuring a diverse range of texts, including technology, sports, society, science, health, games, business or art, among others. The minimum size of a corpus is a matter of discussion. Some authors think that the ideal size, even for general corpus, should be one million words (McEnery & Wilson, 2001, p. 30). However, other authors (Church & Mercer, 1993; Sinclair, 2004) argue that there is no minimum size; the larger the corpus, the better. In this work, I follow these last authors. The enTenTen20 corpus is much larger than other general corpora commonly used such as the COCA corpus[4] (1 billion words) or the BNC corpus[5] (100 million words). All these features guarantee the balance and representativeness of the corpus.

In this study, I use corpus methods to investigate the internal structure of SCIENTIST.[6] Specifically, I conduct two distinct analyses: a collocation analysis (see Hunston, 2002, p. 68; Baker *et al.*, 2013, p. 36) and a KWIC analysis (see Hunston, 2002, p. 39; McEnery & Hardie, 2012, pp. 35-37). Firstly, I examine the most frequent collocates of the expressions "good scientist" and "true scientist." Collocates are the words that occur "frequently within the neighborhood of another word, normally more often than we would expect the two words to appear together because of chance" (Baker *et al.*, 2013, p. 36).

For the collocation analysis, I use a 5:5 span, which signifies that I consider words appearing within five words to the left and five words to the right of the node (i.e., the query term, in this case, "good scientist" and "true scientist"). Additionally, I set a minimum frequency in corpus of five, i.e., only collocates appearing at least five times in the corpus will be included, and a minimum frequency in span of three, i.e., only collocates appearing at least three times within the indicated span will be included. To estimate the statistical significance of the collocates, I used the Mutual Information score (or MI score). The "MI-score indicates the strength of a collocation. [...] In other words, the MI-score measures the amount of non-randomness present when two words co-occur" (Hunston, 2002, p. 71).[7]

By examining collocates, we can discern patterns in the usage of a given term. My focus will be on the top 20 statistically significant collocates. As will be seen in Section 4, considering the top 20 collocates suffices to yield a substantive and comprehensive depiction of the concept's structure. The goal of this collocation analysis is to identify terms that frequently appear with "good scientist" and "true scientist," directly revealing both concrete features and abstract values. For example, one of the most frequent collocates of "true scientist" is "humility," which represents an abstract value. Furthermore, among the top 20 collocations of "good scientist" and "true scientist," some terms may not explicitly denote concrete features or abstract values but still imply them. For instance, "hallmarks" ranks as the fourth most frequent collocate of "good scientist," as seen in phrases like "one of the hallmarks of a good scientist is ..." or "these are the hallmarks of a good scientist: ...," which are common in the corpus. Examining the contexts in which "hallmarks" appears can thus provide insight into the concrete features and abstract values typically associated with "good scientist" and "true scientist."

The KWIC analysis is designed to analyze the contexts of this second type of collocation. A KWIC analysis shows the searched word (or node), for example, "hallmark" or "hallmarks," surrounded by its context. I reviewed all instances of these collocates,[8] and manually extracted the associated concrete features and abstract values. In other words, this part of the analysis is interpretative and entirely qualitative, since the determination of what counts as a concrete feature or as an abstract value depends almost exclusively on my qualitative interpretation of each context.[9] For example, by inspecting the contexts in which "hallmarks" appears as a statistically significant collocation of "good scientist" and

---

4 https://www.english-corpora.org/coca/.

5 https://www.english-corpora.org/bnc/.

6 All the searches were conducted between June 1, 2023, and July 14, 2023.

7 An MI score higher than 3 is considered significant (see Hunston, 2002, p. 71; Baker, 2006, p. 101).

8 For a detailed analysis of the collocations used in the analysis, see Section 4.2.

9 Perhaps it could be thought that the combination of quantitative and qualitative methods is not desirable. However, both methods are not only compatible, but also complementary (see Bordonaba-Plou, 2022).

"true scientist", we can identify different concrete features and abstract values. To illustrate this, consider the next examples:

10) One of the hallmarks of a good scientist is understanding his or her data.
11) Objectivity, scrupulous attention to details, intellectual honesty and dedication to searching for the truth —these are the hallmarks of a good scientist.
12) Erecting hypotheses that can be falsified, and designing experiments capable of doing so, is the hallmark of the true scientist.
13) Humility is the hallmark of any true scientist.

I have selected only those concrete features and abstract values that appear right next to the collocation, i.e., either in the same sentence or another sentence but just before the collocation. For example, if I am analyzing the contexts of "hallmark" as a collocation of "true scientist," and considering 10), then only one abstract value is selected, "humility." In short, the KWIC analysis allows us to extract many of the concrete features and abstract values that competent speakers associate with "good scientist" and "true scientist," simply by investigating the contexts in which these collocations occur.

4. Analysis

As said in the previous section, I will conduct two analyses: a collocation and a KWIC analysis. By combining these two types of analysis, we can construct a comprehensive structure of both the concrete and abstract dimensions of SCIENTIST. In Section 4.1, I will present the results of the collocation analysis for both "good scientist" and "true scientist".[10] Then, in Section 4.2, I will present the results of the KWIC analysis.

4.1. Collocation analysis

For the collocation analysis, I will categorize the collocates into four groups: non-relevant collocates (NR), collocates directly referring to concrete features (CF), collocates directly referring to abstract values (AV), and collocates revealing concrete features and abstract values through their KWIC analysis (KWIC).

Among the top 20 collocates of "good scientist" (see Table 1), there are seven collocates that are not relevant for the analysis. First, "creationist," "atheist" and "incompatible" address the question of whether being a good scientist is compatible with holding a specific religious position. These terms often appear in contexts of uncertainty, and therefore do not assert that a specific religious attitude is a defining feature of scientific practice. Consider the following examples:

14) Can a creationist be good scientist?
15) Did you have to be an atheist to be a good scientist?

Second, "thermodynamics" is used when talking about scientific practice, but not as a required knowledge area for a good scientist, as in:

16) There are gaps in the fossil record, mutations are harmful, or evolution violates the second law of thermodynamics. A good scientist will reject or revise evolutionary theory based on the facts that are presented.

[10] "Good scientist" has an absolute frequency in the corpus of 1,677 occurrences, while "true scientist" has an absolute frequency in the corpus of 890 occurrences.

Third, "Dawkins" and "Hoffman" refer to the famous scientist Richard Dawkins and the scientist and conspiracy theorist Jim Hoffman, respectively. Fourth, "chemist" appears in contexts where the person being discussed happens to be a chemist but where the speaker does not talk of any concrete features or abstract values, as in:

17) He is also a chemist and a good scientist and during one of his experiments his nose evaporates.

Table 1

Top 20 collocates of "good scientist" and "true scientist", with their MI scores and the type of collocate for the analysis

"good scientist"	MI score	Analysis	"true scientist"	MI score	Analysis
skeptic	12.21	AV	Fallacy	14.23	NR
creationist	12.08	NR	spoken	12.77	KWIC
hallmarks	11.67	KWIC	unscientific	12.46	KWIC
disprove	11.62	CF	hallmark	11.21	KWIC
hallmark	10.62	KWIC	humility	10.99	AV
thermodynamics	10.63	NR	fallacy	10.94	NR
communicator	10.49	AV	atheist	10.88	AV
Dawkins	10.35	NR	skeptical	9.92	AV
atheist	10.29	NR	philosopher	9.75	KWIC
scientist	10.17	KWIC	scientist	9.65	KWIC
skeptical	10.17	AV	curiosity	9.45	AV
mathematician	10.18	KWIC	Darwin	9.24	NR
chemist	9.71	NR	researcher	9.08	NR
engineer	9.52	KWIC	discoveries	8.98	CF
incompatible	9.53	NR	Gilbert	8.94	NR
hypothesis	9.50	CF	scholar	8.93	KWIC
philosopher	9.42	KWIC	humble	8.65	AV
qualities	9.23	KWIC	questioning	8.63	AV
like	9.13	KWIC	theories	8.32	KWIC
Hoffman	9.02	NR	knows	8.22	KWIC

Among the collocates directly referring to concrete features, there are two notable examples. First, "disprove" reflects the idea that a good scientist must attempt to disprove her / his own theories. This collocate appears in sentences such as:

18) A good scientist will always try to disprove as well as prove their own theories.

Second, "hypothesis" pertains to a central activity in scientific practice, investigating, holding, testing, or disproving hypothesis. It is used in contexts like:

19) No good scientist holds to any hypothesis despite the evidence.

There are three collocates representing abstract values. On the one hand, "skeptic" and "skeptical" convey the idea that being a good scientist requires adopting a critical attitude towards theories and hypotheses to reach the truth (in this sense, it is similar to "questing for impartial truth"). These terms appear in sentences such as the following:

20) Any good scientist is a skeptic.

21) And a good scientist is always skeptical.

On the other hand, "communicator" suggests that a good scientist must also excel in communication, effectively conveying their research to the layperson. Consider the next example:

22) A good scientist is also a good communicator, who can take the data and explain it clearly to another intelligent and honest seeker of understanding.

Finally, there are eight collocates that reveal the presence of concrete features and abstract values after a KWIC analysis. First, "hallmarks," which appears in constructions such as "one of the hallmarks of a good scientist is ...." Second, "hallmark," as in "The hallmark of a good scientist is ...." Third, "mathematician," found in contexts like "You can't have a good scientist or mathematician without ...." Fourth, "scientist," in sentences such as "And a good scientist is a ... scientist." Fifth, "engineer," as in "You cannot, for instance, be a good scientist or engineer if ...." Sixth, "philosopher," as in "... the philosopher did what a good scientist does, ..." Seventh, "qualities," in phrases like "these are qualities essential for a good scientist." Eighth, "like," as in "like a good scientist, he wanted ...."

In summary, there are seven non-relevant collocates, two that represent concrete features, three that represent abstract values, and eight that, after a KWIC analysis, reveal both concrete features and abstract values.

I will now present the results of the collocation analysis for "true scientist" (see Table 1). Similar to the case of "good scientist," some of the top 20 collocates are not relevant. First, "Fallacy" and "fallacy" refer to the No True Scientist Fallacy, a modern name for Flew's No True Scotsman fallacy. Second, "philosopher" appears in contexts discussing whether someone is both a true scientist and a philosopher, without highlighting any concrete features or abstract values, as in:

23) He is neither a true scientist, nor just a philosopher.

Third, "Darwin" and "Gilbert" refer to the renowned scientists Charles Darwin and William Gilbert, where the speaker simply asserts that they are true scientists. Fourth, "researcher" is used in apposition with "true scientist" and other similar terms such as "truth seeker," which seems to indicate that they are used as mere synonyms.

There is only one collocate indicating concrete features: "discoveries." This suggests that a defining characteristic of a true scientist is the continual pursuit of new discoveries, as in:

24) The true scientist is not a discoverer of this or that; he, or she, embodies a process of ongoing, successive discoveries.

Six different collocations seem to directly indicate the presence of abstract values. First, "questioning" conveys that a hallmark of a true scientist is their continual inquiry, as in 25) and 26):

25) A true scientist never stops questioning.
26) He was a true scientist, always questioning.

I have included this as an abstract value because I consider it to be close, if not similar, to questing for impartial truth. Second, "humility" reflects the notion that a true scientist demonstrates humility in their scientific practice. Examples include:

27) Not all people are able to maintain the humility of a true scientist.
28) He confessed with the humility of a true scientist that he had no explanation.

Third, the use of "atheist" suggests that a true scientist cannot be an atheist, thus indicating the necessity of faith. Unlike in the case of "good scientist,"[11] this term appears in explicit assertions such as:

29) No true scientist could be an atheist.
30) Gruenwald said that no true scientist can be an atheist from a scientific standpoint, it is not possible.

Fourth, "skeptical" refers to the idea that something indispensable to be a true scientist and to reach the truth is to maintain a skeptical attitude, as in:

31) A true scientist is skeptical.
32) I think that a true scientist is always skeptical.

Fifth, "curiosity" highlights that true scientists are driven by curiosity, as in:

33) He has a mind of a true scientist: curiosity.
34) Overall, Sasha is a true scientist with a maddening curiosity that drives his critical analysis and experimentation into new frontiers.

Sixth, "humble" refers to the idea that a true scientist is humble, with examples like:

35) The true scientist remains humble.
36) Of all people in the world, the true scientist is the most humble."[12]

Finally, there are seven collocates that, after a KWIC analysis, indicate concrete features and abstract values related to scientific practice. First, "spoken," as in:

37) As he characteristically told me back then, "we're still in the data-collection stage, no conclusions yet." Spoken like a true scientist.

Second, "unscientific," as in:

38) To Sahelian, however, the failure to present both sides, even when there is only one, is unscientific: "A true scientist takes a fair approach," says Sahelian.

---

[11] This result is not a trivial one. When speakers use "atheist" in connection with "good scientist" they do not use it in association with abstract values, but when they use it in connection with "true scientist" they do. This susggests that the use of DCCs varies depending on context; in other words, that DCCs are context-dependent. In this sense, this finding aligns with other theories that have argued that DCCs are context-dependent. For example, Phillips & Plunkett (2023) denies the purported asymmetry between DCCs and single character concepts. Among other things, they defend that context has a significant influence on whether a concept may elicit dual-character linguistic behavior. In a similar vein, Baumgartner (2024) defends the "pragmatic view" on DCCs, according to which we can find cases where DCCs are used in a totally new and context-dependent way, and where these context-dependent uses can be explained appealing to top-down primary pragmatic processes such as "strengthening" and "loosening" (Recanati, 2004) or ad hoc concept or occasion-specific senses (Carston, 2019).

[12] Note that while some collocates, for example, "humility" or "humble," can be clearly categorized, other collocates do not fit neatly into one of the two categories. For example, "questioning" could be considered a concrete feature since it refers to a specific activity. Similarly, "discoveries" might be seen as an abstract value rather than a concrete feature because it implies an ongoing commitment to scientific inquiry and truth. This suggests that, although the descriptive and normative dimensions are independent, there is an inherent relationship between them.

Third, "hallmark," as in constructions such as "... is the hallmark of any true scientist." Fourth, "scientist," as in "... is not a true scientist. The scientist considers ...." Fifth, "scholar," as in "Any true scientist (or legitimate scholar of any sort) would ...." Sixth, "theories," as in "No true scientist objects to having his theories verified." Seventh, "knows," as in "as every true scientist knows ...."

In summary, there are six non-relevant collocates, one that represents concrete features, six that represent abstract values, and seven which, after a KWIC analysis, reveal both concrete features and abstract values.

4.2. KWIC Analysis

Now, I will introduce the KWIC analysis, beginning with the results of "good scientist". Through meticulous examination of the eight chosen collocates, I could distinguish the following concrete features and abstract values:

— "hallmarks": *concrete features*: questioning concepts, using different viewpoints, understanding the data, and taking risks; *abstract values*: intellectual honesty, questing for impartial truth, and open-mindedness.

— "hallmark": *concrete features*: controlling the ego, and analysis of the applicability of the scientific method; *abstract values*: being guided by the evidence (similar to quest for impartial truth), and trusting in one's convictions.

— "mathematician": *concrete features*: acquittance with art, logical thinking, strategic planning, and using the imagination; *abstract values*: open-mindedness.

— "scientist": *concrete features*: making complicated things seem simple, strategic planning, and not being influenced by politics or money; *abstract values*: curiosity (three occurrences), humility (two occurrences), unquenchable desire to know (similar to quest for impartial truth), modesty, and perseverance.

— "engineer": *concrete features*: wide vision (similar to using different viewpoints), knowledge dissemination, writing and presentation skills, understanding the social context of science, analysis of the circumstances of a problem, disproving own theories, and attention to details; *abstract values*: prescient vision, scientific integrity, faith in the discipline, inner compass, and obsession.

— "philosopher": *abstract values*: being guided by the evidence, and open-mindedness

— "qualities": *concrete features*: logical thinking (three occurrences), strategic planning, exchange of ideas with colleagues and rivals, good verbal reasoning, and strong industrial partnerships; *abstract values*: curiosity (four occurrences), open-mindedness (three occurrences), perseverance (two occurrences), intellectual rigor, being even-handed, being innovative, being collaborative, and integrity.

— "like": *concrete features*: gathering data (six occurrences), conducting experiments (three occurrences), formulating hypothesis (two occurrences), testing hypothesis (two occurrences), considering the professional career from several angles (two occurrences), revising hypothesis in light of new evidence (two occurrences) (similar to test hypotheses), analysis of the applicability of the scientific method, providing references and sources, analysis of the circumstances of a problem, multi-disciplinary approach, ensuring that the product works, testing subjects, attention to details, and identifying a problem; *abstract values*: open-mindedness (five occurrences), being guided by the evidence (four occurrences), searching for the truth, making questions, curiosity, and perseverance.

As can be seen, speakers associate with "good scientist" both concrete features and abstract values. To enhance the coherence of the analysis, I have categorized analogous KWIC findings under unified headings (e.g., "be guided by evidence" or "search for the truth" under the label "quest for impartial truth") and I have aggregated the results of the collocation analysis (e.g., "quest for impartial truth" encompasses occurrences of collocations like "hallmarks" but also from collocations like "skeptic" or "skeptical" (see Table 2)).

Table 2

Concrete features and abstract values related to "good scientist"

CF and "good scientist"	AV and "good scientist"
disproving own theories / hypothesis (9 occurrences)	questing for impartial truth (27 occurrences)
gathering data (6 occurrences)	open-mindedness (11 occurrences)
logical thinking (4 occurrences)	curiosity (8 occurrences)
testing hypothesis (4 occurrences)	good comunicator (5 occurrences)
conducting experiments (3 occurrences)	perseverance (4 occurrences)
strategic planning (3 occurrences)	humility / modesty (3 occurrences)
formulating hypothesis (2 occurrences)	integrity (2 occurrences)
considering the professional career from several angles (2 occurrences)	being collaborative
analysis of the applicability of the scientific method (2 occurrences)	trusting in one's convictions
attention to details (2 occurrences)	prescient vision
analysis of the circumstances of a problem (2 occurrences)	faith in the discipline
using different viewpoints (2 occurrences)	inner compass
taking risks (2 occurrences)	obsession
identifying a problem	intellectual rigor
understanding the data	being even-handed
questioning concepts	being innovative
controling the ego	
not being influenced by politics or money	
acquittance with art	
using the imagination	
making complicated things seem simple	
knowledge dissemination	
writing and presentation skills	
providing references and sources	
good verbal reasoning	
understanding the social context of science	
exchanging ideas with colleagues and rivals	
strong industrial partnerships	
multi-disciplinary approach	
ensuring the product works	
testing subjects	

As before, to determine the concrete features and abstract values mentioned in the contexts of the collocates of "true scientist," I conducted a KWIC analysis. Listed below are the concrete features and abstract values related to each of the seven collocates identified in the collocation analysis:

— "spoken": *concrete features*: gathering data, and application of mathematics.

— "unscientific": *concrete features*: conducting experiments, and considering new hypothesis and evidence; *abstract values*: questing for impartial truth.

— "hallmark": *concrete features*: making hypotheses, and conducting experiments; *abstract values*: humility, curiosity, courage, and love of truth (similar to questing for impartial truth).

— "scientist": *concrete features*: gathering data, seeing the hidden links between facts, using knowledge to help the less-fortunate, and not asking how theories are generated in observation-oriented disciplines; *abstract values*: being an apprentice of the divine, being skeptical, and questing for impartial truth.

— "scholar": *abstract values*: feeling challenged (similar to curiosity), sincerity, love of truth, and open-mindedness.

— "theories": *concrete features*: conducting experiments; *abstract values*: being guided by the evidence, being skeptical, and questing for impartial truth.

— "knows": *concrete features*: gathering data (two occurrences), and conducting experiments; *abstract values*: being skeptical (two occurrences), awareness of the limits of scientific knowledge (two occurrences), being guided by the evidence, the sacredness of nature, humility, and not being overly optimistic.

As evident, the exploration reveals both concrete features and abstract values associated with "true scientist," with a noteworthy difference compared to the findings for "good scientist." In this case, there is a prevalence of abstract values. As before, I have clustered analogous findings under unified headings, and I have included the results of the collocation analysis (see Table 3).

Table 3

Concrete features and abstract values related to "true scientist"

CF and "true scientist"	AV and "true scientist"
conducting experiments (4 occurrences)	questing for impartial truth (19 occurrences)
gathering data (4 occurrences)	humility (13 occurrences)
making new discoveries (2 occurrences)	not being an atheist (4 occurrences)
application of mathematics	awareness of the limits of scientific knowledge (2 occurrences)
considering new hypothesis and evidence	curiosity (2 occurrences)
making hypotheses	courage
seeing the hidden links between facts	being an apprentice of the divine
using knowledge to help the less-fortunate	sincerity
not asking how theories are generated in observation-oriented disciplines	open-mindedness
	the sacredness of nature
	not being overly optimistic

After presenting the results of the analyses, the next section will discuss some ideas directly derived from these findings, elucidating their direct relevance to the discussion on DCCs.

5. Discussion

The results of the analyses show four points warranting discussion regarding the internal structure of DCCs. Firstly, the identification of concrete features and abstract values in the literature appears inadequate in capturing the complexity of SCIENTIST. Secondly, the findings seem to support the idea that there are realizations dependencies between the descriptive and the normative dimension of SCIENTIST. Thirdly, contrary to what is agreed, both the descriptive and the normative dimension interact with "good" and "true," which implies that both expressions can be used to make the two types of normative evaluation proper of DCCs. Lastly, there are some terms that could be used as potential markers for investigating the internal structure of other DCCs.

5.1. The internal structure of scientist: A more elaborate picture

Knobe *et al.* (2013) identified concrete features such as "Conducting experiments, Analyzing data, Developing theories, Writing papers" (Knobe *et al.*, 2013, p. 243), all of which are included in the results. Tasks like gathering and understanding data, along with conducting experiments, could encompass the first two features. Similarly, formulating and testing hypothesis could be interpreted as a facet of theory development. Moreover, there are others that have to do with the idea of writing papers, for example, writing and presentation skills, and providing references and sources. However, other concrete features directly related with them appear in the results, for example, logical thinking, strategic planning, attention to details, analysis of the applicability of the scientific method, or analysis of the circumstances of a problem.

Significantly, the analyses reveal additional concrete features unrelated to the previously mentioned activities, representing different activities of scientific engagement. These features can be delineated into three distinct groups. The first group relates to innovation, which would include the following: considering the professional career from several angles, using different viewpoints, taking risks, acquittance with art, using the imagination, and multi-disciplinary approach. The second group revolves around science communication: making complicated things seem simple, knowledge dissemination, and understanding the social context of science. The third group seem to indicate the interrelation between industry and science: strong industrial partnerships, and ensuring that the product works.

Regarding abstract values, the analyses show that SCIENTIST involves more than one abstract value, "the pursuit of scientific knowledge" (Knobe *et al.*, 2013, p. 243) or "the quest for impartial truth" (Del Pinal & Reuter, 2017, p. 479). While these values are prevalent in the analyses, other abstract values also emerge with notable frequency, for example, open-mindedness, curiosity, perseverance, integrity, humility, or modesty.

In short, a scientist is not just someone who pursues scientific knowledge or impartial truth applying the scientific method, i.e., gathering data, and formulating and testing hypothesis through experiments. Rather, a scientist is characterized by a multifaceted behavior. They demonstrate curiosity taking risks and using the imagination. Moreover, they are committed to making their research accessible to lay audiences, disseminating knowledge and making complicated things seem simple. Furthermore, humility and modesty are hallmarks of their character, as they control the ego and use knowledge to help the less-fortunate.

5.2. Realization dependencies between concrete features and abstract values

The second issue I would like to address pertains to the nature of the relationship between the two dimensions of SCIENTIST and, more notably, what insights it offers for a deeper comprehension of the internal structure of DCCs in a broader context. Knobe *et al.* (2013) defends that the relationship between the two dimensions of DCCs is one of realization, i.e., "the [concrete] features associated with dual character concepts can all be seen as ways of realizing the same abstract values." For them, "Conducting experiments, Analyzing data, Developing theories, Writing papers" (Knobe *et al.*, 2013, p. 243) all serve as avenues to realize one abstract value, "the pursuit of scientific knowledge" (Knobe *et al.*, 2013, p. 243). As Reuter (2019, p. 7) observes, for Knobe and colleagues "the descriptive and normative information seems to be coordinated." I think that the results of the present study could be interpreted as corroborating this judgment, because there are a significant degree of coordination between the diverse concrete features and abstract values derived from the two analyses. On the one hand, the study seems to reproduce the same coordination relation between the descriptive and the normative information as described by Knobe *et al.* (2013). On the other hand, the new concrete features and abstract values that this study identifies could be organized in realizations dependencies similar to those pointed out by Knobe *et al.* (2013) (see Table 4).

Table 4

Realization dependencies between concrete features and abstract values

Concrete features	Abstract values
gathering data	
conducting experiments	questing for impartial truth
making / testing hypothesis	
not being influenced by politics or money	honesty / integrity
using knowledge to help the less-fortunate	
controlling the ego	
not being overly optimistic	humility / modesty
awareness of the limits of scientific knowledge	
making complicated things seem simple	
knowledge dissemination	good communicator
writing and presentation skills	
good verbal reasoning	
considering the professional career from several angles	
using of different viewpoints	open-mindedness
multi-disciplinary approach	
using of imagination	
taking risks	curiosity / being innovative
acquittance with art	

As shown in Table 4, six different realization relationships are derivable from the results. In all of them, possessing specific concrete features serves to realize the associated abstract values. For instance, if I make complicated things seem simple, I am concerned with knowledge dissemination, I have good writing and presentation skills and good verbal reasoning, it could be said that I am a good communicator. In short, there appears to be a significant coordination between the various concrete features and abstract values that speakers refer in using "scientist." Given that the study reveals new concrete features and abstract values related to the DCC SCIENTIST, one might argue that there is not a single realization relationship, as Knobe *et al.* (2013) suggest, but rather several. It is important to emphasize that the six different realization relationships proposed in this study represent just one possible way of organizing the vast amount of data obtained in the analysis. Whether this phenomenon is unique to this concept or prevalent across other DCCs warrants empirical investigation. Exploring other DCCs, e.g., ARTIST or SOLDIER, and founding that their internal structures are composed of multiple and distinct concrete features and abstract values, and that there are distinct coordination relations between them, could bring us closer to establishing this as a common trait among all DCCs.

To conclude this second point, it is worth briefly considering the possibility that a single concrete feature may realize different abstract values. Consider, for example, the notion of making new discoveries. Is this activity or concrete feature one that manifests the abstract value of the quest for impartial truth, or does it equally reflect the abstract value of open-mindedness? Depending on the chosen criterion, a given concrete feature will realize one abstract value or another. Further-

more, some results —such as "questioning," which I mentioned earlier— can be classified either as concrete features or as abstract values. All of this supports the earlier point that there are different ways of organizing the data. However, there are limitations on the realization relationships between concrete features and abstract values. For example, gathering data or conducting experiments does not seem to realize the abstract values of curiosity or being a good communicator. Likewise, the use of imagination or familiarity with art does not appear to realize the abstract value of seeking impartial truth.

5.3. The interactions of "good" and "true" with the descriptive and the normative dimensions

The third point that merits discussion has to do with the fact that both search expressions, "good scientist" and "true scientist," are linked both to concrete features and abstract values. Specifically, when searching for "good scientist," the results show nearly an equal number of concrete features and abstract values. As illustrated in Table 2, there are 65 concrete features and 69 abstract values. However, when searching for "true scientist," the results indicate a higher number of abstract values compared to concrete features. As Table 3 shows, there are 16 concrete features and 46 abstract values. This implies that the expression "good scientist" does not predominantly highlight the descriptive dimension over the normative dimension, nor does the expression "true scientist" *exclusively* highlight the normative dimension. I emphasize "exclusively" because there are many cases where speakers straightforwardly associate concrete features with the term "true scientist." Consider the following examples:

39) Erecting hypotheses that can be falsified, and designing experiments capable of doing so, is the hallmark of the true scientist.

40) As he characteristically told me back then, "we're still in the data-collection stage, no conclusions yet." Spoken like a true scientist.

41) Any true scientist (or legitimate scholar of any sort) would consult an advanced scientific text for definitions of technical terms, especially when attempting to criticize them.

42) The true scientist is not a discoverer of this or that; he, or she, embodies a process of ongoing, successive discoveries.

43) Any true scientist tries to rigorously validate the analogy before using it as a component of theory.

In short, speakers use both expressions to indicate that the person they refer to as a scientist possesses certain concrete features or that their behavior conforms to certain abstract values. This finding is significant because the literature suggests that "good" is more directly associated with the descriptive dimension, while "true" is more closely linked to the normative dimension. However, the results of this study suggest that, contrary to what is agreed, both the descriptive and the normative dimension interact with "good" and "true." Note that this does not imply that the two dimensions are not independent. It remains possible to classify someone as a scientist if s/he possesses some of the concrete features but does not strive to achieve any of the abstract values usually associated with the concept. Similarly, it is still possible to classify someone as a scientist when the opposite is the case.

At this point, we can ask ourselves: what is the significance of this finding for other studies on DCCs? I believe this finding is highly impactful, as it is widely assumed that the "true"-modifier highlights the normative dimension of DCCs. However, this study reveals that, while the normative dimension interacts with "true," it also interacts with "good." In fact, the opposite is also true, the descriptive dimension interacts both with "good" and "true." This suggests that, contrary to what is generally accepted, both expressions can be used to convey the two types of normative evaluation characteristic of DCCs. Consequently, these results challenge most existing theories about the "true"-modifier. Further studies will be needed to determine whether this pattern holds for other DCCs. It is possible that this is a unique characteristic of SCIENTIST and that, for instance, speakers do not associate (or rarely associate) concrete features with "true artist," just as they do not associate abstract values with "good artist."

5.4. Potential markers for investigating the internal structure of DCCs

The collocation analysis revealed several terms that, although they do not directly refer to concrete features or abstract values, have proven effective for investigating the internal structure of SCIENTIST when examined in context. These terms are: "hallmarks," "hallmark," "qualities," and "knows." Another reason that makes them suitable for being employed in future studies is their significant frequency. When searching for "good scientist," "knows" appears 46 times, "qualities" 15 times and "hallmark(s)" 11 times. Similarly, when searching for "true scientist," "knows" appears 16 times, and "hallmark" four times. Their high frequency is due to their occurrence in diverse constructions that are commonly used. Consider the following expressions:

— one of the hallmarks of a good scientist

— the hallmark of + a / the / any + good / true + scientist

— qualities + of / for / needed to be / make for + a good scientist

— a / any / every / the + good / true + scientist knows

These four constructions (or variations of them) could be used in future studies to investigate the internal structure of other DCCs. In this way, we could investigate whether the results obtained in the present study are replicated for other DCCs, specifically, whether the expressions "good + DCC" and "true + DCC" both interact with the descriptive and the normative dimensions of DCCs. A search in the enTenTen20 corpus in Sketch engine shows that some of the terms appearing with SCIENTIST also appear with other DCCs.[13] For example, "qualities" is the fourth most statistically significant collocate of "good teacher," the tenth of "good mentor," the 16th of "true friend," and the 41st of "true teacher"; "knows" is the 24th most statistically significant collocate of "true musician," the 30th of "good musician," the 30th of "good minister," and the 48th of "good teacher"; "hallmark"[14] is the 13th most statistically significant collocate of "good mentor," and the 17th of "good theory"; finally, "hallmarks" is the 22nd most statistically significant collocate of "good theory."

To sum up, there are several terms that, in addition to being used frequently, will allow us to extract concrete features and abstract values of DDCs through an analysis of their contexts. Two groups of terms are particularly noteworthy. The first group includes "hallmark(s)," "mark(s)" and "qualities," which refer to the characteristics of a good and true scientist. The second group includes "knows," which relate to what a good and true scientist knows. A thorough investigation of the internal structure of other DCCs will involve a detailed analysis of their most frequent specific collocates, which may, of course, vary from those listed here. However, these terms provide a solid starting point to systematize the search and analysis involved in the complex task of investigating the internal structure of DCCs.

6. Conclusions

The study of DCCs is still in its early stages, as it has been less than a decade since these concepts were first identified. However, research on DCCs is rapidly increasing, both in the number of studies and in the diversity of philosophical areas in which they are applied. In this article, I have explored one of the most under-examined topics in the literature: the internal structure of DCCs.

---

[13] The DCCs used in the search were extracted from the Appendix A of Knobe *et al.* (2013).

[14] There are also other collocates directly related to "hallmarks," for example, "mark" is the 41st most statistically significant collocate of "good minister," and "marks" is the 48th most statistically significant collocate of "true musician."

I have focused on SCIENTIST, one of the most representative examples of DCCs. Using corpus methods, I have been able to investigate the two ways speakers use for characterizing their instances. I have searched for the expressions "good scientist" and "true scientist" in the enTenTen20 corpus in Sketch Engine, as these expressions, according to the literature, interact more directly with the descriptive and normative dimension, respectively. The analyses revealed four key findings. Firstly, the internal structure of SCIENTIST is significantly more complex and nuanced than previously reported. Regarding the descriptive dimension, it includes not only typical tasks such as gathering data, conducting experiments, developing theories, and writing scientific articles but also includes concrete features related to innovation and science communication. Concerning the normative dimension, it incorporates the quest for impartial truth, as highlighted in the literature, and other abstract values such as open-mindedness, curiosity, perseverance, integrity, humility, or modesty.

Secondly, the realization relationship between the descriptive and the normative dimension highlighted in Knobe *et al.* (2013) extends beyond the concrete characteristics and the abstract value they mentioned, encompassing other concrete features and abstract values.

Thirdly, the findings challenge most existing theories about the "true"-modifier, as it is widely assumed that the "true"-modifier highlights the normative dimension of DCCs. However, this study reveals that both "good" and "true" interact with both the descriptive and the normative dimension. Searching for the expression "good scientist" yields nearly equal numbers of concrete features and abstract values. In contrast, searching for "true scientist" yields more abstract values than concrete features, but there are numerous instances where "true scientist" is used in connection with concrete features. This suggests that, contrary to common assumptions, both expressions can convey the two types of normative evaluation characteristic of DCCs.

Fourthly, several terms appearing in the analysis, for example, "hallmark(s)," "mark(s)," "qualities," and "knows" could be useful for investigating the internal structure of other DCCs.

To conclude, I want to stress the importance of conducting further studies on the internal structure of DCCs. Without additional research, we cannot fully understand the extent of the findings from this study. For instance, is it plausible to assume that the internal structure of DCCs like ARTIST, SOLDIER or JAZZ is similarly complex or akin to that of SCIENTIST? Until we investigate whether speakers associate concrete features or abstract values when using expressions such as "true artist," "good soldier," or "true jazz," it is challenging to generalize the findings of the present study.

Acknowledgemets

This work has been carried out within the project *Intuiciones y Filosofía Experimental del Lenguaje* (PID2023-150396OA-I00), funded by the Ministerio de Ciencia, Innovación y Universidades, the Agencia Estatal de Investigación and the FEDER. I express my gratitude to Joshua Knobe, and the audience of the *Conference on Dual Character Concepts, Normative Generics and Evaluative Language*, held in Zurich from June 16, 2022 to June 18, 2022, for their valuable comments on an earlier version of this paper. I would also like to thank the two anonymous reviewers from *Theoria* who evaluated my work as their comments have led to significant improvements.

REFERENCES

Alfano, M., Higgins, A., & Levernier, J. (2018). Identifying virtues and values through obituary data-mining. *The Journal of Value Inquiry, 52*, 59-79. https://doi.org/10.1007/s10790-017-9602-0

Almeida, G., Struchiner, N., & Hannikainen, I. (2023). Rule is a dual character concept. *Cognition, 230*, 105259. https://doi.org/10.1016/j.cognition.2022.105259

Andow, J. (2015a). How "intuition" exploded. *Metaphilosophy, 46*(2), 189–212. https://doi.org/10.1111/meta.12127

Andow, J. (2015b). How distinctive is philosophers' intuition talk? *Metaphilosophy, 46*(4-5), 515-538. https://doi.org/10.1111/meta.12151

Ashton, Z., & Mizrahi, M. (2018). Intuition talk is not methodologically cheap: Empirically testing the "received wisdom" about armchair philosophy. *Erkenntnis, 83*, 595-612. https://doi.org/10.1007/s10670-017-9904-4

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Baumgartner, L. (2024). The pragmatic view on dual character concepts and expressions. *Mind and Language, 39*(5), 726-744. https://doi.org/10.1111/mila.12505

Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media studies: The representation of Islam in the British press*. Cambridge University Press.

Baz, A. (2012). *When words are called for: A defense of ordinary language philosophy*. Harvard University Press.

Bluhm, R. (2013). Don't ask, look! Linguistic corpora as a tool for conceptual analysis. In M. Hoeltje, T. Spitzley, & W. Spohn (Eds.), *Was dürfen wir glauben? Was sollen wir tun? Sektionsbeiträge des achten internationalen Kongresses der Gesellschaft für Analytische Philosophie e.V.* (pp. 7-15). DuEPublico.

Bluhm, R. (2016). Corpus analysis in philosophy. In M. Hinton (Ed.), *Evidence, experiment and argument in linguistics and philosophy of language* (pp. 91-109). Peter Lang.

Bordonaba-Plou, D. (2021). An analysis of the centrality of intuition talk in the discussion on taste disagreements. *Filozofia Nauki, 29*(2), 133-156. https://doi.org/10.14394/filnau.2021.0008

Bordonaba-Plou, D. (2022). Bucles virtuosos: Una defensa del uso de métodos cualitativos y cuantitativos en filosofía del lenguaje. *Alpha. Revista de Letras, Artes y Filosofía, 54*, 59-70.

Carston, R. (2019). Ad hoc concepts, polysemy and the lexicon. In K. Scott, B. Clark, & R. Carston (Eds.), *Relevance, pragmatics and interpretation* (pp. 150–162). Cambridge University Press.

Caton, J. N. (2020). Using linguistic corpora as a philosophical tool. *Metaphilosophy, 51*(1), 51-70. https://doi.org/10.1111/meta.12405

Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics, 19*(1), 1-24.

Dancy, J. (1995). In defence of thick concepts. *Midwest Studies in Philosophy, 20*, 263-279. https://doi.org/10.5840/msp19952016

Del Pinal, G., & Reuter, K. (2017). Dual character concepts in social cognition: Commitments and the normative dimension of conceptual representation. *Cognitive Science, 41*(S3), 477-501. https://doi.org/10.1111/cogs.12456

Evans, V. (2009). Lexical Concepts. In V. Evans (Ed.), *How Words Mean: Lexical Concepts, Cognitive Models, and Meaning Construction* (pp. 127-148). Oxford University Press.

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Frege, G. (1891/1997). Function and concept. In M. Beaney (Ed.), *The Frege reader* (pp. 130-148). Blackwell Publishing.

Frege, G. (1892/1997). On Sinn and Bedeutung. In M. Beaney (Ed.), *The Frege reader* (pp. 151-171). Blackwell Publishing.

Guo, C., Dweck, C. S., & Markman, E. M. (2021). Gender categories as dual-character concepts? *Cognitive Science, 45*(5), e12954. https://doi.org/10.1111/cogs.12954

Flanagan, B., & Hannikainen, I. R. (2022). The folk concept of law: Law is intrinsically moral. *Australasian Journal of Philosophy, 100*(1), 165-179. https://doi.org/10.1080/00048402.2020.1833953

Hansen, N., Porter, J. D., & Francis, K. (2021). A corpus study of "know": On the verification of philosophers' frequency claims about language. *Episteme, 18*(2), 242-268. https://doi.org/10.1017/epi.2019.15

Hinton, M. (2021). Corpus linguistics methods in the study of (meta)argumentation. *Argumentation, 35*, 435-455. https://doi.org/10.1007/s10503-020-09533-z

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference CL* (pp. 125-127).

Kenny, A. (2010). Concepts, brains, and behaviour. *Grazer Philosophische Studien, 81*(1), 105-113. https://doi.org/10.1163/9789042030190_007

Kirchin, S. (2013). Introduction: Thick and thin concepts. In S. Kirchin (Ed.), *Thick concepts* (pp. 1-19). Oxford University Press.

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition, 127*, 242-257. https://doi.org/10.1016/j.cognition.2013.01.005

Knobe, J. (2022). Personal identity and dual character concepts. In K. Tobia (Ed.), *Experimental philosophy of identity and the self* (pp. 49-70). Bloomsbury.

Leslie, S. J. (2015). "Hillary Clinton is the only man in the Obama administration": Dual character concepts, generics and gender. *Analytic Philosophy, 56*(2), 111-141. https://doi.org/10.1111/phib.12063

Liao, S., Meskin, A., & Knobe, J. (2020). Dual character art concepts. *Pacific Philosophical Quarterly, 101*(1), 102-128. https://doi.org/10.1111/papq.12301

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Mejía-Ramos, J. P., Alcock, L., Lew, K., Rago, P., Sangwin, C., & Inglis, M. (2019). Using corpus linguistics to investigate mathematical explanation. In E. Fischer & M. Curtis (Eds.), *Methodological advances in experimental philosophy* (pp. 239-263). Bloomsbury.

Phillips, J., & Plunkett, D. (2023). Are there really any dual-character concepts? *Philosophical Perspectives, 37*(1), 340-369. https://doi.org/10.1111/phpe.12194

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

Reuter, K. (2019). Dual character concepts. *Philosophy Compass, 14*(1), e12557. https://doi.org/10.1111/phc3.12557

Schwitzgebel, E., Huang, L. T., Higgins, A., & Gonzalez-Cabrera, I. (2018). The insularity of Anglophone philosophy: Quantitative analyses. *Philosophical Papers, 48*(1), 21-48. https://doi.org/10.1080/05568641.2018.1429741

Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Sytsma, J., Bluhm, R., Willemsen, P., & Reuter, K. (2019). Causation attributions and corpus analysis. In E. Fischer & M. Curtis (Eds.), *Methodological advances in experimental philosophy* (pp. 209-238). Bloomsbury.

Tobia, K. P., Newman, G. E., & Knobe, J. (2020). Water is and is not H2O. *Mind & Language, 35*(2), 183-208. https://doi.org/10.1111/mila.12234

Väyrynen, P. (2011). Thick concepts and variability. *Philosopher's Imprint, 11*(1), 1-17.

Väyrynen, P. (2013). *The lewd, the rude and the nasty*. Oxford University Press.

Williams, B. (1985). *Ethics and the limits of philosophy*. Routledge.

Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell.

David Bordonaba-Plou is an Assistant Professor at the Universidad Complutense de Madrid. His research interests center on experimental philosophy of language, political philosophy of language, digital humanities and the role of intuitions.

Address: Universidad Complutense de Madrid, Facultad de Filosofía, Dept. de Lógica y Filosofía Teórica, Despacho B-09, Edif. B, Calle del Prof. Aranguren, s/n, Ciudad Universitaria, 28040, Madrid, Spain. E-mail: davbordo@ucm.es – ORCID: https://orcid.org/0000-0002-0788-9733

Theoria, 2025, 40(1), 68-80
https://doi.org/10.1387/theoria.26482

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Against Pritchard's refutation of epistemic relativism

(Contra la refutación de Pritchard del relativismo epistémico)

Matías Gariazzo*
Universidad de la República
https://orcid.org/0000-0002-6346-8386

Keywords

Epistemic relativism
Deep disagreement
Rationality
Duncan Pritchard

ABSTRACT: This paper reconstructs Duncan Pritchard's (2011, pp. 278-284; 2021, pp. 1120-1122; 2025, pp. 56-58) refutation of epistemic relativism and presents an objection to it. This refutation presupposes that epistemic relativism would be true in case there were rationally irresolvable deep disagreements. Pritchard's refutation, thus, amounts to an argument purporting to show that all deep disagreements are rationally resolvable. Our objection, in turn, aims to show that the examples of rationally resolvable deep disagreement Pritchard presents have particular features that, while making them rationally resolvable, not all deep disagreement has. In order for these examples to be representative of all deep disagreements we need to accept a particularly strong notion of rationality. Pritchard's (2011) notion of a *truth-seeker* presupposes a strong notion of rationality that could play that role. In recent papers, in contrast, Pritchard (2023, pp. 305-308; 2025, p. 53) makes use of a weaker rationality notion in characterizing deep disagreements. Both these alternatives prove to be problematic for Pritchard's refutation. On the one hand, if the notion of rationality used to characterize deep disagreements secures their rational resolvability, it will already presuppose the falsehood of epistemic relativism. On the other hand, if the refutation treads on a weaker rationality notion, it will simply fail to give reasons to think that all deep disagreements can be rationally resolved. Be that as it may, we claim that Pritchard's work allows us to identify a subset of deep disagreements that have a particular structure that makes them rationally resolvable.

Palabras clave

Relativismo epistémico
Desacuerdo profundo
Racionalidad
Duncan Pritchard

RESUMEN: Este artículo reconstruye la refutación del relativismo epistémico formulada por Duncan Pritchard (2011, pp. 278-284; 2021, pp. 1120-1122; 2025, pp. 56-58) y presenta una objeción a ella. Esta refutación presupone que el relativismo epistémico sería verdadero si hubiese desacuerdos profundos racionalmente irresolubles. Así, la refutación de Pritchard equivale a un argumento que pretende mostrar que todos los desacuerdos profundos son racionalmente resolubles. Nuestra objeción, por su parte, busca mostrar que los ejemplos de desacuerdo profundo racionalmente resoluble presentados por Pritchard tienen rasgos particulares que, si bien los hacen racionalmente resolubles, no todo desacuerdo profundo tiene. Para que estos ejemplos sean representativos de todo desacuerdo profundo es preciso manejar una noción particularmente fuerte de racionalidad. La noción de buscador de la verdad de Pritchard (2011) presupone una noción de racionalidad fuerte que podría cumplir este rol. En artículos recientes, en cambio, Pritchard (2023, pp. 305-308; 2025, p. 53) maneja una noción más débil de racionalidad al caracterizar los desacuerdos profundos. Ambas alternativas resultan problemáticas para la refutación de Pritchard. Por un lado, si la noción de racionalidad usada en la caracterización de los desacuerdos profundos asegura su resolubilidad racional, ella ya presupondrá la falsedad del relativismo epistémico. Por el otro, si la refutación maneja una noción de racionalidad más débil, ella no nos ofrecerá razones para pensar que todo desacuerdo profundo puede ser racionalmente resuelto. Sea como fuere, sostenemos que el trabajo de Pritchard permite identificar un subconjunto de desacuerdos profundos que tienen una estructura particular que los hace racionalmente resolubles.

* ***Correspondence to:*** Matías Gariazzo. Instituto de Filosofía, Facultad de Humanidades y Ciencias de la Educación, Universidad de la República. 268 Reconquista, Montevideo, Uruguay – mgariazzo1@gmail.com – https://orcid.org/0000-0002-6346-8386

How to cite: Gariazzo, Matías (2025). «Against Pritchard's refutation of epistemic relativism»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 68-80. (https://doi.org/10.1387/theoria.26482).

Received: 10 June, 2024; Final version: 03 December, 2024.

ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

1. Introduction

As its title indicates, this paper presents an objection to Duncan Pritchard's (2011, pp. 278-284; 2021, pp. 1120-1122; 2025, pp. 56-58) refutation of epistemic relativism. Both this objection and this refutation assume that epistemic relativism would be true in case there were rationally irresolvable deep disagreements. Thus, following Pritchard, we shall address the question of the rational resolvability of deep disagreements, understanding the categories of epistemic relativism, deep disagreement and rational resolvability in ways that make this question directly relevant for the discussion over the truth of epistemic relativism.

For Pritchard (2011, p. 269; 2021, pp. 117-118), deep disagreements are philosophically interesting because they suggest the existence of the phenomenon of epistemic incommensurability and, as a consequence, the truth of epistemic relativism. He (2011, p. 269) defines epistemic incommensurability as follows:

> *Epistemic incommensurability*:
>
> It is possible for two agents to have opposing *propositional acceptances* which are rationally justified to an equal extent where there is no rational basis by which either agent could properly persuade the other to revise their view.[1]

Epistemic relativism, in turn, is understood here —following Pritchard (2011, pp. 269-270; 2021, 1117-1118)— as the thesis that the rational justification of propositions and propositional attitudes is relative to the system of commitments (specially including our acceptances of propositions) of an agent, where different such systems can give rise to epistemic incommensurability.[2] Thus, the existence of epistemic incommensurability entails the truth of epistemic relativism.

We can get an even better grip of this understanding of epistemic relativism by distinguishing it from truth relativism. According to Pritchard (2009, pp. 406-411), the only plausible form of epistemic relativism is one that understands the relativized notion of rational justification in a purely internalist way. This is compatible with holding that the epistemic standing of our beliefs or convictions has an externalist aspect that should not be relativized to a system of commitments. For instance, we can take the truth of such a belief or conviction as a necessary condition for it to count as knowledge, which is clearly a positive epistemic standing, while understanding truth in non-relativist and non-internalist terms as correspondence to some state of affairs. An analogous thing can be said about the notion of reliability as applied to beliefs and other propositional attitudes, which has been used by externalist authors to give accounts of knowledge. The kind of epistemic relativism that Pritchard has in mind is not one that rejects these externalist notions, but simply one that relativizes a purely internalist notion of rational justification. Pritchard (2009, p. 407) called this sort of epistemic relativism *dialectical epistemic relativism*, and considered that it did not pose significant philosophical problems. Be that as it may, this form of relativism does constitute a substantial view about the limits of rationality and ra-

---

1 In his paper from 2011 Pritchard talks of beliefs and not of propositional acceptances when defining epistemic incommensurability. However, based on his later work, there is reason to think that he would define this notion using a more general propositional attitude, like simply accepting or being committed to a proposition. More precisely, Pritchard (2016, pp. 90-94) thinks that certain propositions that give rise to deep disagreements cannot be believed, insofar as they cannot be justified, and often (Pritchard, 2011, p. 268, p. 280, pp. 282-284; 2016; 2021, p. 1119) uses the term "conviction" for the positive attitude we can have towards them. Be that as it may, a conviction, just as a belief, would be a way of accepting a proposition. This is the reason why we have substituted "beliefs" for "propositional acceptances" in the previous definition.

2 It is worth noting that in a recent paper Pritchard (2025, p. 56) understands the expression "epistemic relativism" differently. According to this latter understanding, epistemic relativism is just the thesis that there are different systems of commitments relative to which people can make rational evaluations, no matter whether these systems are or can be epistemically incommensurable (i.e. whether they lead or can lead to incompatible evaluations). Accordingly, the argument that we here consider as a refutation of epistemic relativism is presented there by Pritchard (2025, pp. 56-58) just as a refutation of epistemic incommensurability.

tional discussion, and in later works Pritchard (2011, pp. 269-270; 2021, 1117-1118) simply refers to it as epistemic relativism and tries to refute it by showing that we have reason to take deep disagreements as rationally resolvable.

In order to see the question of the truth of (dialectical) epistemic relativism as analogous to the question of whether deep disagreements are rationally resolvable, we also need to accept the following (at least) necessary condition for deep disagreements:

> *Rational persistence*:
>
> A disagreement is rationally persistent if and only if it is resistant to rational resolution without the parties being irrational in holding their respective views.

Notice that in the previous definition we do not assume that deep disagreements are rationally irresolvable, but we do suggest that, in case they were rationally resolvable, they resolution would be particularly difficult. As Chris Ranalli (2021, p. 985) points out, we should not take deep disagreements as rationally irresolvable by definition because the question of whether the disagreements that are usually seen as deep are rationally resolvable has no obvious answer and, as a result, we should leave it open in order to obtain a better understanding of disagreements and their rationality. And, in case there could be both rationally resolvable and rationally irresolvable deep disagreements, this question should be addressed in connection with each particular disagreement that we classify as deep.

Finally, we understand the rational resolvability of a disagreement in the following way:

> *Rational resolvability:*
>
> The disagreement between A and B over proposition *p* is rationally resolvable if and only if there is an attitude *D* which is the only rational attitude that A and B can have towards *p*.[3]

According to these understandings of epistemic relativism, deep disagreement and rational resolution, if there were rationally irresolvable (deep) disagreements, epistemic relativism would be true. Hence, in order to refute epistemic relativism, we need to show or at least give good reasons to think that all deep disagreements can be rationally resolved, since the mere existence of a subset of rationally irresolvable deep disagreements would result in the truth of epistemic relativism. This is what Pritchard (2011, pp. 278-284; 2021, pp. 1120-1122; 2025, pp. 56-58) tries to do. He offers reasons to think that all deep disagreements are rationally resolvable that are, at the same time, reasons to think that epistemic relativism is false. As we have pointed out, the purpose of this paper is to question these reasons, and so to defend epistemic relativism from Pritchard's attack.

In the next and second section, we present Pritchard's (2011, pp. 278-284; 2021, pp. 1120-1122; 2025, pp. 56-58) refutation of epistemic relativism. In order to do so, we introduce and analyze some key features of his view (2011, 2021, 2023, 2025) and present a case of a (rationally resolvable) deep disagreement in order to see how Pritchard's procedure for rational resolution works. This latter case is a hypothetical reconstruction of the disagreement that Galileo and Cardinal Bellarmine had over the truth of geocentrism, where Bellarmine could —if he lived enough time— rationally change his view so to agree with Galileo. This reconstruction is based on Pritchard's (2011, pp. 280-282) description of this case, but also on some features that he (Pritchard, 2023, pp. 305-308, pp. 312-313; 2025, pp. 51-53) currently thinks deep disagreements have. In the third section we first argue that this case, as a description of a rationally resolvable deep disagreement, must have certain particular structural features that are not, in principle, generalizable to all deep disagreements. In order for an example like this one to be representative of all deep disagreements we need to accept a particularly strong notion of rationality, and read the above-mentioned principles in terms of it. Pritchard's

[3] Ranalli (2021, pp. 4977-4978) distinguishes different senses of rational resolvability. The definition in the main text corresponds to one of such senses.

(2011) notion of a *truth-seeker* presupposes a strong notion of rationality that could play that role. In recent papers, in contrast, Pritchard (2023, pp. 305-308; 2025, p. 53) makes use of a weaker notion of rationality in characterizing deep disagreements. We end this section by arguing that both these alternatives are problematic for Pritchard's refutation of epistemic relativism. On the one hand, if the notion of rationality used to characterize deep disagreements secures their rational resolvability, it will already presuppose the falsehood of epistemic relativism, and so this refutation can be accused of begging the question. On the other hand, if the refutation treads on a weaker rationality notion, it will fail to give reasons to think that all deep disagreements are susceptible of rational resolution. Be that as it may, we shall claim that Pritchard's work allows us to identify a subset of deep disagreements that have a particular structure that makes them rationally resolvable. In the fourth and final section we summarize the main points made in this paper.

2. Pritchard's refutation

Pritchard's (2011, pp. 278-284; 2021, pp. 1120-1122; 2025, pp. 56-58) refutation of epistemic relativism is made from an epistemological view that is based on Ludwig Wittgenstein's *On Certainty* (1969). Initially, Pritchard (2011, 2021) arguably saw all deep disagreements as being, directly or indirectly, over a hinge proposition. Hinge propositions are the ones that, in a given context, are held with the maximal level of certainty and cannot be doubted. Consequently, they give the framework that makes our justification practices and meaningful doubts possible. According to Pritchard (2021, p. 1119), an agent can put forward a proposition as a reason to hold or doubt another proposition if and only if she is more certain of the former than the latter. Thus, insofar as hinge propositions are held with the maximal level of certainty, they cannot be justified and cannot be *directly* —without, as we shall see, an indirect procedure that presupposes a change of context— criticized. Since Pritchard (2016, pp. 90-94) understands belief as a propositional attitude that, by definition, is capable of being justified, he thinks that hinge propositions cannot be believed. Be that as it may, they would be held in another way, and the pro attitudes we would have towards them —which are often called "convictions" by Pritchard (2011, p. 268, p. 280, pp. 282-284; 2016, 2021, p. 119)— could give rise to deep disagreements.[4] Pritchard (2023, 2025) has recently introduced some changes in his view on deep disagreements. According to his current understanding of this category (Pritchard, 2023, pp. 302-309; 2025, pp. 51-54, pp. 58-59), not all deep disagreements are hinge disagreements, and there could even be hinge disagreements that are not deep. In a nutshell, he now sees deep disagreements as genuine disagreements that concern, directly or indirectly, an axiological question of deep existential importance to the parties involved, where these parties are broadly rational, are truly committed to their opposing judgments and have considered these judgements (i.e. they are not tacit).[5] Accordingly, deep hinge disagreements are disagreements over an axiological hinge proposition, but not all axiological propositions that give rise to deep disagreements are hinges. Despite this change in his view, Pritchard (2025, pp. 56-58) still gives the same reasons as he gave before (Pritchard, 2011, pp. 278-284; 2021, pp. 1120-1122) to think that deep hinge disagreements are rationally resolvable, and, as we shall briefly see, what happens in the case of these disagreements could be taken as a reason to

[4] The examples of hinge propositions found in the literature are varied, and go from ones that are usually taken to be cornerstones of our way of thinking and presuppositions of many regular propositions (Wright 2004, pp. 188-194; 2014, pp. 214-217; Coliva and Palmira, 2021, pp. 408-409), like "There are physical objects" and "There are other minds beside my own", to ones that are about very specific aspects of someone's context, like "I have never been to the Moon" or "I have two hands" (Wittgenstein, 1969, § 133, § 245, § 247, § 252, § 667). As we shall see, Pritchard (2011, pp. 283; 2021, p. 1121; 2023, pp. 308-310; 2025, pp 54-55.) takes the latter propositions as hinges, and recently (2023, pp. 308-313; 2025, pp. 55-56) came to think that there are basic axiological propositions that should also be considered as hinges, and that are the ones that can give rise to deep hinge disagreements. It is worth noting that, without rejecting the admittedly vague characterization of hinge propositions given in the main text, some authors have given more precise definitions that exclude some of these examples from the set of hinges. In particular, Wright (2004, pp. 188-194; 2014, pp. 214-217) and Coliva and Palmira (2021, pp. 408-409) give definitions that exclude propositions that are about specific aspect of someone's context.

[5] Other properties that are usually assigned to deep disagreements, like their resistance to rational resolution, are meant to follow from this characterization.

doubt that there are rationally irresolvable deep disagreements simpliciter. Thus, in what follows we shall focus only on the case of deep hinge disagreements, bearing in mind that, on Pritchard's current view, they are axiological in nature.

The previously sketched view seems to entail the rational irresolvability of deep hinge disagreements. The rational resolution of a disagreement is something that can be achieved by means of at least one of the disagreeing parties' revision of the relevant propositional attitude in light of reasons or evidence. But insofar as our commitments to hinge propositions cannot be justified, they seem incapable of being revised in such a way. However, Pritchard (2011, pp. 278-284; 2021, pp. 1119-1122; 2025, pp. 56-58) holds that deep hinge disagreements are rationally resolvable by appealing to reasons or evidence, though in an indirect way. Such a resolution would occur by the rational revision, carried out by at least one of the parties to the disagreement, of attitudes (typically belief or disbelief) towards certain propositions that are not hinges, revision that would have as a consequence the revision of the attitude had towards the hinge proposition(s) that the disagreement concerns. As a result, there would be no case for epistemic relativism.

Pritchard (2011, p. 283; 2021, p. 1121) gives examples of how I could rationally change my attitudes towards the propositions "I have never been to the Moon" and "I have two hands," so that they lose their hinge status and could, eventually, be denied. Regarding the first proposition, he (2011, p. 283; 2021, p. 1121) invites us to imagine a future where space travels become very frequent, so that I could have passed by the Moon without remembering it. According to this example, I would, in the first place, have rationally revised a range of beliefs concerning the world I live in, as this world itself would have changed over time, and in the second place, this revision would have the consequence that the proposition that I have never been to the Moon would lose its hinge status and would become possible to deny it. Concerning the second proposition, Pritchard (2021, p. 1121) asks us to imagine ourselves waking up in the hospital after a car accident, or (Pritchard, 2011, p. 283) being stumbling, bewildered and confused from the wreckage of a plane crash. In such a context, immediately after taking notice of the situation one is in, one could reasonably doubt that one has two hands, and accordingly this proposition would lose its hinge status and denying it would become a possibility.

Be that as it may, how could this procedure of rational change have an impact on the resolution of deep hinge disagreements? After all, no disagreement with one previous self is involved in these examples, and not only does Pritchard (2021) not claim that there is, but he (Pritchard, 2023, pp. 309-310; 2025, pp. 54-56) even claims that there cannot be deep disagreements over propositions such as "I have never been to the Moon" or "I have two hands," insofar as their truth or falsehood is something every rational person can agree on. Regarding the absence of a disagreement with one previous self in these examples, notice that Pritchard (2011, p. 283; 2021, p. 1121) treats the just mentioned propositions as time-neutral or tensed propositions. Otherwise, when in these examples the context shifts, the agent will not be doubting or denying the same proposition that she accepted before. In other words, the agent would be just doubting or denying a new proposition about a particular time while retaining her acceptance of another proposition about another previous time. But do these cases involve disagreements with one previous self once we treat these propositions as being time-neutral? Well, as John MacFarlane (2007, pp. 22-23; 2014, pp. 130-133) shows, two different evaluations of one time-neutral proposition respectively made at two different times, or a pair of acceptances of two time-neutral contradictory propositions respectively made at different times, do not give rise to a disagreement, at least in an intuitive sense of disagreement that could motivate any form of relativism. Suppose that, in the morning, John accepts the time-neutral proposition that Mary is in her bed, while, at night, Ann rejects this proposition and accepts the time-neutral proposition that Mary is not in her bed. For MacFarlane, John and Ann's acceptances and eventually assertions should be assessed for accuracy (i.e. correctness in a propositional truth derived sense) considering the time of each one's respective acceptance. And since these times are different, their acceptances can both be accurate, which means that they are not in conflict.[6] Thus, MacFarlane shows that, according to our intuitive sense of disagreement (or at least the one involved

[6] The position MacFarlane (2007, pp. 22-23; 2014) puts forward as reasonable for time-neutral or tensed proposition is a nonindexical contextualist one, according to which the time that is relevant for assessing for accuracy an assertion or acceptance of such a proposition is the time of the context of the assertion or acceptance. He (2014, pp. 44-70) also contends that there are propositions that are neutral with respect to other things, like a standard of taste, and that assertions or acceptances of such propositions are to be assessed for accuracy consid-

in philosophical discussions over relativism), one person disagrees with another if and only if they have two propositional acceptances that cannot both be accurate. In case we countenanced the existence of tensed propositions, this would typically happen if (i) the parties respectively accept or assert two propositions that are inconsistent with each other and (ii) their acceptances/assertions must be assessed for accuracy relative to the same time and possible world (i.e. the same circumstance of evaluation).[7] For our purposes, the upshot of all this is that the previous cases are not examples of rational change involving a disagreement with one's previous self, but just examples of the indirect procedure of rational change that, in cases of deep disagreement, would make rational resolution possible. In order to find an example of deep hinge disagreement and see how it could be rationally resolved we arguably need to countenance other types of hinge propositions in our hinge epistemology.

As we pointed out, Pritchard (2023, pp. 302-308; 2025, pp. 54-56) now thinks that the hinge propositions that can give rise to deep disagreements are axiological in nature. According to him, disagreements involving a clash between a religious and a secular world-picture are among the typical cases of deep disagreement (Pritchard, 2023, pp. 301-305; 2025, pp. 51-53) and, in particular, of deep hinge disagreement (Pritchard, 2021, p. 1119; 2023, pp. 312-313; 2025, pp. 55-56). Pritchard (2023, 2025) does not give us examples of the particular hinge commitments that would be involved in the latter disagreements. Be that as it may, in a passage where he (Pritchard, 2023, p. 313) is talking about religious hinge commitments he states that "basic religious conviction might be encapsulated in a creed, for example, where it is part of the religious practices to publicly assert these claims." Insofar as the content of such religious commitments can be asserted and are part of a creed, they could include, for instance, that the world was created by God or that the Bible is a holy book containing truths about the world. Both contents seem to be factual and to have epistemic significance, since they would condition what we take as legitimate evidence for propositions over certain subject matters. Be that as it may, for Pritchard both of them would also be axiological, insofar as they shape a way of experiencing the world and valuing different ways of living.[8] Pritchard (2023, p. 302, p. 304) also makes clear that deep disagreements can indirectly concern an axiological claim, while directly and on the surface concerning a purely factual claim. In this regard, he states the following about purely scientific debates that he considers expressions of deep disagreements:

> Consider the kinds of large-scale scientific debates that look like plausible candidates to be deep disagreements, such as debates about whether our scientific understanding of agency allows for free will, whether science can accommodate qualia, or debates about the origin of the universe. While such discussions do not concern axiological claims of the relevant kind directly, they certainty concern them indirectly. If we lack free will, for example, then that threatens to make our lives absurd. Or consider a dispute about the origin of the universe and the ramifications this might have for whether one embraces or rejects a religious worldview. (Pritchard, 2023, p. 304)

In light of these considerations, we can re-describe a case of rationally resolvable deep disagreement presented by Pritchard (2011, pp. 280-282) as a disagreement that indirectly concerns an axiological hinge proposition.[9] This reading seems reasonable, insofar as the case involves a clash between a religious and a secular world-picture. The case is the

---

ering, say, the standard of taste relevant at the context of the assessor, not the one relevant at the context of the person whose assertion or acceptance is being evaluated (of course, these contexts coincide when someone assesses her own assertions or acceptances).

7 It is worth noting that we can think of different scenarios that have the upshot that two acceptances/assertions cannot be jointly accurate. Suppose that John asserts now that Mary is in her bed, that Mary asserted 24 hours ago that Mary will not be in her bed in 24 hours, and that the asserted propositions are time-neutral. These propositions would not be inconsistent in the sense of being impossible for them to be jointly true at one single circumstance of evaluation composed by a particular world and time. Still, Johns and Mary's acceptances/assertions could not be jointly accurate.

8 In the same vein, Pritchard (2023, p. 310) claims that propositions like "There is an external world" or "The future resembles the past," despite stating objective facts and not making explicit reference to our sources of knowledge, work as epistemic principles.

9 Pritchard (2011, pp. 278-284) actually presents two cases of allegedly rationally resolvable deep disagreement. Both of them involve a clash between a religious and a secular world-picture, and so can be re-described —in line with Pritchard's (2023, pp. 302-308; 2025, pp. 51-53) current view— as being about an axiological proposition.

disagreement that Galileo and Cardinal Bellarmine would have had at the beginning of the seventeenth century over whether heliocentrism (according to which the sun is at the center of the universe and the earth orbits around it) or geocentricism (according to which the earth is at the center of the universe and the sun orbits around it) is true. According to Pritchard's (2011, pp. 280-282) description of this disagreement, Bellarmine would have grounded his geocentric view on a literal interpretation of the Bible taken as a holy book, whereas Galileo would have grounded his heliocentric view on empirical observations made with his telescopes. Thus, we can, for our purposes, take this disagreement as concerning, indirectly, a hinge proposition held by Bellarmine such as "The Bible, as a holy book, provides evidence for the location of heavenly bodies," and see —in line with Pritchard's current view— this proposition as axiological as well as epistemic.[10, 11] Whereas Bellarmine would have accepted this proposition, Galileo would have rejected it. On the other hand, on Pritchard's description of the case Bellarmine would have been open to the possibility of being refuted and persuaded to change his mind by means of the presentation of empirical scientific evidence against his view. It would just have happened that such a robust evidence was not available at the time. Recall that Galileo's telescopes were quite rudimentary and that arguably, at that time, there was not conclusive evidence for heliocentrism. According to this, if the empirical evidence we now have had been presented to Bellarmine, he would have changed his mind and admitted that heliocentrism was true. This would have happened by means of the same sort of indirect procedure that Pritchard (2011, p. 283; 2021, p. 1121) claims that allows someone to change her commitments to hinge propositions like "I have two hands" or "I have never been to the Moon." First, by means of taking knowledge of such evidence, Bellarmine would have incorporated several beliefs concerning empirical observations. These beliefs would have, as its contents, non-hinge propositions. Second, this belief change would have forced Bellarmine to deny the proposition that the Bible, as a holy book, provides evidence for the location of heavenly bodies, which was previously held by him as a hinge.[12]

We shall consider this case in more detail in the next section. For now, we need to see what reasons Pritchard gives to hold that all deep disagreements can be rationally resolved by such an indirect procedure. After all, even if we grant to him that the previous disagreement is both deep and rationally resolvable, it could happen that it has particular features that not all deep disagreements have.

According to Pritchard (2011, pp. 282-283; 2021, pp. 1120-1121; 2025, pp. 57-58) there is a universally held hinge commitment that creates the strong presumption that there always are shared resources (beliefs and rational procedures) between any two subjects that allow them to rationally resolve any disagreement they could have.[13] This universal hinge commitment, which Pritchard (2021, pp. 1120-1123) calls *über hinge commitment*, states that we (oneself and others) are not radically and systematically in error in our propositional attitudes. Hinge propositions that are not universal would be expressions, in particular contexts, of this universally held hinge commitment. In a given context, doubting our most entrenched commitments about the world we inhabit (e.g. the conviction that I have never been to the Moon), would cast doubt over the über hinge commitment, and this —as we shall briefly see— would undermine communication and, as a result, make language impossible. We need a change of context, which would go hand in hand with a change in some of our beliefs in regular propositions, for a doubt in a particular hinge proposition to be possible. In Pritchard's own words:

---

10 To be sure, in order to evaluate Pritchard's view on the rational resolution of deep disagreements we need not take a stance on whether our reconstruction of this disagreement is a true description of the disagreement that Galileo and Bellarmine actually had. Rather, we should treat the described case as a *possible* case of rationally resolvable deep disagreement, and see whether the conclusions we can draw about it can be extrapolated to any other deep disagreement.

11 Pritchard (2011, pp. 280-282) does not explicitly say which the hinge proposition involved in the previous case is. We chose to interpret the case in a way that is in line with his current view on deep hinge disagreements.

12 To be sure, a disagreement —and, in particular, a deep disagreement— could be rationally resolved by means of the refutation of both opposing views or by showing that they are equally and inconclusively supported by the available evidence. For the sake of simplicity, we shall here only focus on cases where it can be shown that one of the disagreeing parties is wrong and the other is right.

13 Pritchard (2011, 2016, 2021) does not describe the content of this commitment as a proposition. Accordingly, he only talks of a universal hinge commitment but not of a universal hinge proposition.

> [...] if one can change a subject's wider beliefs to a sufficient extent, then this will have a bearing on which claims function as manifestations of the über hinge commitment. This thus explains how even deep hinge disagreements can be rationally resolved via a kind of 'side-on' persuasion, whereby one doesn't target the hinge commitments directly, but rather the opponent's wider set of beliefs. Since it is these beliefs that ensure that a hinge commitment to this specific proposition is a manifestation of the overreaching über hinge commitment, it follows that if one can effect significant change in these beliefs, then one can also change an opponent's hinge commitments. (Pritchard, 2025, p. 58)

Pritchard (2011, p. 282; 2021, p. 1122) holds that the über hinge commitment must be accepted in order to make sense of the behavior of others and ourselves. Thus, without this universal commitment, communication and, as a result, language itself would be impossible. More precisely, based on Donald Davidson's (1986, pp. 314-319) work on radical interpretation, Pritchard claims that one's understanding of what someone says, and so one's attribution of fine-grained beliefs to her, requires that we apply the principle of charity when interpreting her. This principle commands one, as the only means to interpret another person's utterances, to assume that the majority of her propositional acceptances are, by one's lights, true, and so the same as one's own. Only by means of this procedure one can, eventually, identify the differences that may exist between our propositional acceptances and this other person's. In other words, what one considers mistakes can only be intelligible with the help of a background of propositional acceptances that are shared with the person one is interpreting. Accordingly, we cannot make sense of the idea of there not being a significant number of propositions (including some non-universal hinge propositions) that are jointly accepted by any two people we can understand. And, in particular, we cannot make sense of the idea of there not being such common resources between two people we see as having a deep disagreement. Thus, the point Pritchard makes is that, once we accept the existence of such common resources, we must grant that any deep disagreement can, in principle, be rationally resolved by making use of them.[14]

To be sure, in case it is correct, Pritchard's argument from the impossibility of someone being massively mistaken does not amount to a conclusive proof that all deep hinge disagreement can be rationally resolved by means of such common resources. Rather, this would be a reasonable presumption to make once we admit that we cannot make sense of someone being mistaken in such a way. And this presumption would result in someone who holds that not every deep disagreement is rationally resolvable having the burden of proof. Moreover, since on Pritchard's (2023, pp. 302-308; 2025, pp. 58-59) present view, not all deep disagreements are hinge disagreements, his current case against epistemic relativism must include the claim that the result found in the case of deep hinge disagreements is reason to think that all deep disagreements can be rationally resolved. In this vein, considering the possibility of there being deep disagreements that do not concern hinges, Pritchard (2021) briefly states the following:

> [...] there is the possibility that there is a sub-class of deep disagreements which don't involve hinge commitments, and hence concerns fundamental commitments that aren't held in an essentially arational manner. But in that case, why would we think that deep disagreements involving these commitments would lead to epistemic incommensurability and thus epistemic relativism anyway? At the very least, we are owed an argument as to why this might be so, and the prospects for such an argument do not look promising, given what we have seen with regard to how this difficulty is meant to play out in the supposedly more problematic case of a Wittgensteinian hinge epistemology. (Pritchard, 2021, p. 1124)

[14] Referring to a dispute motivated by a deep hinge disagreement, Pritchard claims: "while one might resort to persuasion rather than reason in order to resolve the dispute [...] there will be an appropriate epistemic path to resolution available since such disputes inevitably occur relatively to a shared background of commitments". (Pritchard, 2011, pp. 282-283)

We will not address this claim here. Our objection, which shall be presented in the next section, questions Pritchard's argument for the rational resolvability of all deep hinge disagreement. And if this objection is correct, there is no room for making the previous claim.

3. An objection to Pritchard's refutation

In order to show why I think the previous argument is unconvincing, it is useful to analyze in more detail the disagreement between Galileo and Bellarmine as we, based on Pritchard (2011, pp. 280-282), have described it in order to see it as both deep and amenable to rational resolution by means of the indirect procedure introduced in the previous section.

As we suggested, this disagreement would be rationally resolvable insofar as Bellarmine was open to revise his view in case there was sufficiently convincing empirical evidence in favor of Galileo's position, and this evidence can be obtained despite the fact that it was absent during Bellarmine's lifetime. We could describe the conditions that makes this disagreement rationally resolvable in favor of Galileo as follows:

(i) Bellarmine accepts the proposition that the Bible, as a holy book, is a source of evidence of the location of heavenly bodies, which for him has the status of a hinge.

(ii) Bellarmine accepts, as Galileo does, an explanation principle that results in his openness to empirical evidence in favor of heliocentrism, that has for him the status of a hinge and that he puts before other considerations when choosing between two alternative explanations of an empirical fact. Such a principle may be, for instance, a principle of greater predictive power.[15]

(iii) Bellarmine recognizes and is disposed to recognize the same empirical facts as Galileo and contemporary astronomers do.

(iv) These empirical facts, assuming an explanation principle like the one described in (ii), speaks in favor of heliocentrism and against geocentrism.

Thus, in this disagreement, Bellarmine accepts two hinge propositions that, while not contradictory, conflict with each other once new information is presented to him. This information is not picked up by his initial system of commitments, so there would be no incoherence in this system. On the other hand, in this disagreement Bellarmine is committed beforehand to privilege one of these hinge propositions, namely the relevant explanatory principle, in case such a conflict emerged. The first point means that, while lacking the relevant new information, Bellarmine is not violating this explanatory principle in accepting that the Bible is a source of evidence of the location of heavenly bodies, while the second point means that he is committed to privilege the first proposition over the second if a conflict emerged.[16]

---

[15] Bear in mind that the predictive power of an explanation (or theory) can be assessed not only with respect to future facts, but also with respect to past facts that have already been verified. One just needs to consider whether these past facts were or could be predicted by means of the explanation and other even earlier facts.

[16] One could wonder whether the proposition mentioned in (i) should be taken as a hinge once we consider that this explanation principle should override it in case of conflict. But, despite this circumstance, the former proposition has the traits that Pritchard (2021, p. 118-119; 2023, pp. 308-309: 2025, p. 54) thinks that characterize hinges: in Bellarmine's initial context, this proposition determines what counts as a reason for a given view about the location of heavenly bodies and cannot be doubted. Notice that, on Pritchard's view, the situation would be similar to what happens to a proposition like 'I have two hands' once a relevant change of context occurs. Presumably, I am disposed to doubt this proposition in relevant new contexts because I am not willing to reject a hinge proposition like 'I have not been systematically hallucinating the recent events of my life.' My acceptance of this latter proposition allows me to incorporate new beliefs about my surroundings that ultimately make the former proposition lose its hinge status. Notice also that 'I have not been systematically hallucinating the recent events of my life' is not equivalent to the über hinge commitment, since one can imagine a context in which it can be

According to this, the possible rational resolution of a deep disagreement is indirect. In a first instance, and as a result of acquiring new information, the party that can be refuted would have to add new beliefs in non-hinge propositions and probably discard others he had. And in a second instance, this modification would give rise to a conflict between the hinge proposition(s) the disagreement concerns and one or more other hinges that are responsible for this person's openness to this new information, conflict that, insofar as this person is committed to privilege the latter hinges in case of such a conflict, would lead her to reject the hinge proposition(s) that the disagreement concerns.[17]

According to this, rationally resolvable deep disagreements have the following structure. The party that can be refuted, despite not being incoherent or guilty of irrationality, accepts two sets of hinge propositions that can conflict with each other if she adds new obtainable information to her belief system. One set contains the hinge propositions the disagreement concerns, while the other contains the propositions that explain her openness to this new information.[18] Finally, this party is committed beforehand to privilege the latter propositions in case such a conflict emerged. Conditions (i) to (iv) guarantee us that the disagreement between Galileo and Bellarmine, as we described it, has this structure. Other conditions would play this role in the case of other deep disagreements. We can grant Pritchard that disagreements that have the structure just described can be seen as being both deep and rationally resolvable. But do we have reason to think that all deep disagreements have this structure?

As we saw, for Pritchard (2011, pp. 269-270; 2021, 1117-1118; 2025, pp. 56-58) there would be a reasonable presumption that there are always common resources to rationally resolve a disagreement. The impossibility of making sense of massive error forces us to admit that there are common resources between any two speakers we find intelligible, and this admission would place the burden of proof on those who argue that not every deep disagreement is rationally resolvable. But the impossibility of massive error and the existence of common resources does not give rise to this presumption. This is so because we can make a significant distinction between those common resources that are needed to secure understanding and those common resources that are needed to rationally resolve a deep disagreement. As a matter of fact, Davidson (1986, pp. 316-318) shows that, in order for two people who do not initially share a language to be able to understand each other, they need to share many occasional beliefs (i.e. beliefs about observable events) (Davidson 1986, p. 316, p. 318) and most beliefs about basic logical questions (Davidson 1986, p. 316). It is an agreement on this sort of beliefs that guarantees understanding and makes language possible. Briefly put, this would happen by means of the radical interpreter's identification of observable events that cause certain assertions of the speaker, allowing her to assign meanings to the asserted sentences that are about such events. To be sure, the principle of charity also commands the radical interpreter to assume that the speaker agrees with her about other sort of issues (i.e. issues that are not about observable events or basic logical questions), but with respect to these issues this principle works as a guide that may lead this interpreter to identify significant differences she has with the speaker. And convictions about such issues are typically involved in the examples of deep disagreement found in the literature (e.g. disagreements about the law of abortion). Hence, nothing in Davidson`s view makes us think that there will always be common resources to rationally resolve a deep disagreement.

As a matter of fact, there is a reasonable presumption in favor of the impossibility to rationally resolve many deep disagreements. Think about the hypothetical situation where Galileo and Bellarmine had the disagreement we have described, but where the second or third condition we have identified for the rational resolution of this deep disagreement was absent. It is difficult if not impossible to see how, in such a scenario, they could rationally resolve their disagreement, but it is still possible to see them —accepting Davidson's view— as two people that can understand each

doubted without granting that one could be systematically mistaken about almost everything (e.g. a situation where one is aware of recently having taken a hallucinatory drug).

17 As we pointed out, a disagreement —and, in particular, a deep disagreement— could be rationally resolved by means of the refutation of both opposing views or by showing that they are equally and inconclusively supported by the evidence. To device such a different case we must modify the fourth condition just presented for the rational resolution of the disagreement between Galileo and Bellarmine.

18 To be sure, both these sets can contain, like in our case study of deep disagreement, only one proposition.

other. According to this, the possibility of devising hypothetical cases where it seems impossible to rationally resolve a deep disagreement, gives rise to the reasonable presumption that many deep disagreements cannot be so resolved. Thus, deep disagreements that could be rationally resolve are, in principle, only those that have the structure we identified a few paragraphs back and that our case study of deep disagreement has in virtue of satisfying conditions (i) to (iv).

It is worth pointing out that Pritchard (2011) presumably makes use of a notion of rationality in light of which it is plausible that any deep hinge disagreement is rationally resolvable. He (Pritchard, 2011, pp. 267-269) claims that in order for a disagreement to be genuinely deep, the parties must be *truth-seekers* with respect to the subject matter of the disagreement. A truth-seeker is a person who has a disposition to change her view in light of proper reasons and evidence. But since our hinges determine what we see as *proper* reasons or evidence for a thesis, this mere disposition does not guarantee the rational resolution of a deep hinge disagreement. In order to guarantee that, then, truth-seekers must share enough hinges to be able, at least in the long run, to agree on how to assess reasons and evidence. According to Pritchard's (2011, p. 268) description of a truth-seeker, she is someone who responds to reasons and empirical evidence more or less in the same way as the current scientific community does. That is to say, she would have roughly the same criteria as this community has to assess an argument or a piece of evidence. And, insofar as one of the parties to a disagreement is not a truth-seeker, his behavior with respect to the subject matter the disagreement is about will not be rational and, as a consequence, the disagreement will not be deep. Recall that a disagreement's rational persistence (at least a necessary feature of a deep disagreement) was characterized as its resistance to rational resolution without the parties being irrational in holding their respective views. If the irresolvability of a disagreement were due to the mere irrationality of one of the parties, the disagreement would be of a trivial non-philosophically interesting sort.

Once our notion of rationality includes the condition of being a truth-seeker, this notion might ensure that a hinge disagreement that is truly deep satisfies conditions that, like (i) to (iv) in the case of the deep hinge disagreement we have considered, make it rationally resolvable.[19] Here we have used, in line with most of the literature on this subject, a notion of rationality that does not demand the fulfillment of such conditions. Consequently, we have held that there is good reason to think that many deep disagreements are rationally irresolvable. We will not defend here our less demanding notion of rationality. It suffices to say that to seriously consider the possibility of epistemic relativism, Pritchard's refutation should not use, from the start, a notion of rationality that guarantees that all deep hinge disagreements are rationally resolvable. Otherwise, the refutation will presuppose the falsity of epistemic relativism.

In recent papers, Pritchard (2023, pp. 305-308; 2025, p. 53) makes use of a weaker notion of rationality in characterizing deep disagreements. He states, as a necessary condition for a deep disagreement that the parties must be broadly rational (2023, p. 311), or that they must be minimally rational (2025, p. 53). Such a requirement is meant to exclude cases where at least one of the disagreeing parties is guilty of sheer incoherence (e.g. holds some beliefs that are inconsistent with a hinge she endorses, or claims to reject a hinge but her belief system presupposes that she accepts it), stubbornness or dogmatism. But, as we have seen, merely being open to change one's view in light of good reasons or evidence cannot secure the rational resolution of a deep hinge disagreement, since our hinges determine what we see as *proper* reasons or evidence for a thesis. Moreover, when thinking about hypothetical cases, we cannot see how a deep hinge disagreement can be rationally resolved unless conditions that guarantee their rational resolvability —like (i) to

[19] Despite the initial plausibility of the thesis that Pritchard's (2011) notion of rationality secures the rational resolvability of all deep disagreements, there is reason to doubt it. Pritchard's (2011, pp. 267-269) description of a truth-seeker refers to how a person assesses empirical evidence and arguments about matters of fact. If there were axiological hinge propositions that have no implications whatsoever for which factual beliefs someone has, there would be deep hinge disagreements that could not be resolved by gathering empirical evidence and putting forward arguments about matters of fact. In such cases, Pritchard's indirect procedure should work, instead, by adding purely axiological information (whatever that may be) that would result in a belief change that would ultimately lead at least one of the parties to rationally change one or more of these axiological hinge commitments with no factual implications. Notice that we should have reason to consider this belief change as rational in order to see the whole process as rational. These considerations show how strong a notion of rationality securing the rational resolvability of all deep hinge disagreements may need to be.

(iv) in the case of the deep disagreement we considered— are satisfied. And if our notion of rationality is not strong enough to secure that conditions of this sort are satisfied for all genuinely deep hinge disagreements, the reasonable presumption to make is that many of them are rationally irresolvable. Despite this, Pritchard (2025, pp. 56-58) still claims that we have reason to think that all deep hinge disagreements are rationally resolvable by means of his indirect procedure for rational change. In this section I gave reasons to think that this claim is mistaken.

In sum, Pritchard has two alternatives, both of which are problematic for his refutation of epistemic relativism. On the one hand, he can opt to use in his characterization of deep disagreements a strong notion of rationality that secures that all deep hinge disagreements are rationally resolvable. If he does this, this notion already presupposes the falsehood of epistemic relativism, and so his refutation can be accused of begging the question. This is specially the case because he (2011; 2016; 2021; 2023; 2025) does not argue for the acceptance of such a notion of rationality. On the other hand, if he makes use of a weaker notion, his argumentation simply fails to give reasons to think that all deep disagreements are susceptible of rational resolution and so that epistemic relativism is false.

Be that as it may, we have also shown that Pritchard's work has a positive result: it allows us to identify a subset of deep disagreements that have a particular structure that makes them rationally resolvable. The four conditions put forward at the beginning of this section are an attempt to identify this structure in the case of the deep disagreement between Galileo and Bellarmine, as we reconstructed it.

4. Final remarks

I have leveled an objection to Pritchard's refutation of epistemic relativism that consists in questioning that all deep disagreements satisfy conditions that, like (i) to (iv) in the case of the deep disagreement we have considered, make them rationally resolvable by means of an indirect procedure. A sufficiently strong notion of rationality could guarantee that conditions of this sort are satisfied by all deep hinge disagreements. Pritchard's refutation, then, faces a dilemma. On the one hand, if such a notion is accepted without argument in a refutation of epistemic relativism, the argument begs the question against this view. On the other hand, if Pritchard's refutation is red in terms of a weaker notion of rationality, like the one he uses in later papers (2023, pp. 305-308; 2025, p. 53) to characterize deep disagreements, it simply fails to give reasons to think that all deep disagreements are amenable to rational resolution and so that epistemic relativism is false.

Despite our conclusion concerning Pritchard's attempt to refute epistemic relativism, we have also claimed that his argumentation has a positive result, namely singling out an indirect procedure of rational change that, coupled with a reasonable characterization of deep disagreement, allows us to identify a set of disagreements that are both deep and rationally resolvable. A deep disagreement needs to satisfy conditions that make it have a particular structure in order to be rationally resolvable, and we have granted that it is possible for these conditions to be satisfied in particular cases -even if it were extremely difficult to find real cases satisfying them. But insofar as we do not think that, in the disagreement between Galileo and Bellarmine, the second and third conditions should be taken as partially clarifying the notion of rationality used in the definitions of epistemic relativism, rational persistence and rational resolvability, we do not consider Pritchard's refutation of epistemic relativism as successful. In other words, we do not think there is good reason to treat the examples of rationally resolvable deep disagreement as representative of all deep disagreements. Rather, these examples would be of a very particular sort, namely examples of disagreement were conditions like the ones laid down in the previous section are met. Be that as it may, as long as we presented an example of rationally resolvable deep disagreement and considered it as a possible hypothetical case, we see Pritchard's refutation as having the positive result just mentioned.

To be sure, as we pointed out, both Pritchard's refutation of epistemic relativism and my objection to it, assume particular understandings of epistemic relativism, deep disagreement and rational resolvability. These understandings,

which were made explicit by means of definitions presented at the beginning of the paper, are what makes the question of the truth of epistemic relativism directly related to the question of the rational resolvability of deep disagreements.

Acknowledgments

I am grateful to CSIC, Universidad de la República, for the funds they gave me for taking part in a research project on deep disagreement, their characterization and possibility of rational resolution.

REFERENCES

Coliva, A. and Palmira, M. (2021). Disagreement unhinged, constitutivism style. *Metaphilosophy*, 52 (3-4): 402-415.

Davidson, D. (1986). A coherence theory of truth and knowledge. In E. Lepore (ed.), *Truth and interpretation: Perspectives on the philosophy of Donald Davidson* (pp. 307-319). Cambridge: Blackwell.

MacFarlane, J. (2007) Relativism and disagreement. *Philosophical Studies*, 132, 17-31.

MacFarlane, J. (2014) *Assessment sensitivity*: *Relative truth and its applications*. Oxford: Clarendon Press.

Pritchard, D. (2009). Defusing epistemic relativism. *Synthese*, 166 (2), 397-412.

Pritchard, D. (2011). Epistemic relativism, epistemic incommensurability, and Wittgensteinian epistemology. In S. D. Hales (ed.), *A companion to relativism* (pp. 266-285). Oxford: Wiley-Blackwell.

Pritchard, D. (2016). *Epistemic angst: Radical skepticism and the groundlessness of our believing*. Princeton, New Jersey: Princeton University Press.

Pritchard, D. (2021). Wittgensteinian hinge epistemology and deep disagreement. *Topoi*, 40 (5), 1117-1125.

Pritchard, D. (2023). Understanding deep disagreement. *International Journal of Philosophical Studies*, 31 (3), 301-317.

Pritchard, D. (2025). Deep disagreement. In M. Baghramian, J. A. Carter and R. Cosker-Rowland (eds.), *The Routledge handbook of philosophy of disagreement*. London: Routledge.

Ranalli, C. (2021). What is deep disagreement? *Topoi*, 40 (5), 983-988.

Wittgenstein, L. (1969). *On certainty*. G. E. M. Anscombe y G. E. von Wright (eds.), D. Paul y G. E. M. Anscombe (trads.). Oxford: Blackwell.

Wright, C. (2004). On epistemic entitlement: Warrant for nothing (and foundations for free). *Aristotelian Society*, Supplementary vol. 78 (1), 167-212.

Wright, C. (2014). On epistemic entitlement (II): Welfare state epistemology. In: D. Dodd and E. Zardini (eds.), *Scepticism and perceptual justification* (pp. 213-247). Oxford: Oxford University Press.

Matías Gariazzo holds a PhD in Philosophy from the University of London and is a professor at the University of the Republic (Uruguay). He has worked mainly on topics belonging to philosophy of language, epistemology and argumentation theory.

Address: Instituto de Filosofía, Facultad de Humanidades y Ciencias de la Educación, Universidad de la República. 268 Reconquista, Montevideo, Uruguay. E-mail: mgariazzo1@gmail.com – ORCID: https://orcid.org/0000-0002-6346-8386

Theoria, 2025, 40(1), 81-97
https://doi.org/10.1387/theoria.26121

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

The unfeasibility of onto-representationalism

(La inviabilidad del onto-representacionalismo)

Mariano Martín-Villuendas*
University of Salamanca
https://orcid.org/0000-0002-6814-7346

Keywords

Veritism
Representationalism
Realism
Models
Idealization

ABSTRACT: Much effort has been devoted to explaining in what sense models represent their corresponding target systems. This has been considered a pivotal philosophical problem since representational models have been widely assumed to canalize our knowledge and understanding of reality. The aim of the paper is to analytically structure the framework commonly adopted to address the Scientific Representation Problem (SR-P), i.e., onto-representationalism, and to examine its main problems. Due to its very theoretical conditions, I conclude that onto-representationalism constitutes an inadequate meta-scientific platform to approach SR-P. I locate the problem in the semantic assumption. To materialize these analyses, I examine the main arguments proposed by the main variants of onto-representationalism: classical onto-representationalism and sophisticated onto-representationalism.

Palabras clave

Veritismo
Representacionalismo
Realismo
Modelos
Idealización

RESUMEN: Se ha dedicado un esfuerzo notable a intentar explicar cómo los sistemas modelos representan a sus correspondientes sistemas objetivo. Dado que los modelos representacionales son centrales a la hora de canalizar nuestro conocimiento y comprensión de la realidad, se ha considerado que el Problema de la Representación Científica (P-RC) constituye una cuestión filosófica central. El objetivo del presente artículo es estructurar analíticamente el marco comúnmente adoptado a la hora de abordar el P-RC, i.e., el onto-representacionalismo, y examinar sus principales problemas. Argumento que el onto-representacionalismo constituye una plataforma meta-científica inadecuada a la hora de ofrecer una solución satisfactoria al P-RC. Localizo las dificultades en el presupuesto semántico. Para materializar estos análisis, examino los principales argumentos propuestos por las principales variantes del onto-representacionalismo: el onto-representacionalismo clásico y el onto-representacionalismo sofisticado.

1. Introduction

Currently, philosophers of science widely accept that non-epistemic factors such as ethical values, economic, or political incentives play a pivotal role in the institution and development of scientific practices. However, much less attention has been paid to analyzing the important theoretical role and implications of meta-scientific assumptions on scientific practice. One author who appreciated their significance was J. H. Woodger (1929). In his doctoral thesis, published as *Biological Principles*, he demonstrated how scientific practice relies on elements that are not strictly scientific. What is more, he consistently argued that they were the origin of many of the everlasting debates that pervade the biology of his time —e.g., Mechanism vs. Vitalism, Organism vs. Environment, or Structure vs. Function. Hereafter, by

* ***Correspondence to:*** Mariano Martín-Villuendas. Department of Philosophy, Logic, and Aesthetics – Instituto Universitario de Estudios de la Ciencia y la Tecnología, University of Salamanca. Edificio I+D+I, Calle Espejo n.º 2, 37007, Salamanca, Spain – marianomv@usal.es – https://orcid.org/0000-0002-6814-7346

How to cite: Martín-Villuendas, Mariano (2025). «The unfeasibility of onto-representationalism»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 81-97. (https://doi.org/10.1387/theoria.26121).

Received: 07 March, 2024; Final version: 26 November, 2024.

ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

meta-scientific assumptions I mean the set of theoretical assumptions —ontological, epistemological, and semantic— used to define how scientific work has proceeded and/or how it should proceed.

In this paper, I intend to revive Woodger's fundamental intuition in light of a current debate within the general philosophy of science: the problem of scientific representation. Certainly, this debate has taken place primarily in the philosophical sphere. In the mid-1960s, philosophers of science realized that scientists, instead of studying phenomena directly, usually rely on simpler hypothetical systems to reason about the phenomenon of interest. Far from being a mere heuristic complement to theories, scientific models seemed to play an essential role in the scientific process of understanding the world. This shift towards "a model-based science" (Godfrey-Smith, 2006) significantly altered the contours of the philosophy of science, giving rise to a series of ongoing debates that resonate to this day: How should we understand the relationships between models and those entities, phenomena, or portions of the world that the models represent?; In order to understand a phenomenon, is it necessary that our models accurately reflect all aspects of interest or only the causally central ones?; What is the role of idealizations and abstractions? I suggest locating this apparently heterogeneous set of questions in a more global and central problem, namely "the Scientific Representation Problem" (SR-P):

> *SR-P*: In what sense do our model systems represent their corresponding target systems and allow scientists to gain knowledge and understanding of reality?

Echoing Woodger, this debate, although philosophical in nature, can eventually shape and impact real scientific practice. If we think of science as just another human activity, it seems hard to deny that scientists do make use of certain meta-scientific assumptions in their daily scientific practice: when interpreting the construction, manipulation and cognitive salience of models and representations. They are not ideal agents isolated from all those beliefs deemed "non-scientific". Consequently, nothing prevents scientists from being influenced by the generally accepted and widely shared theoretical view on scientific representation. Although SR-P is a fundamentally conceptual puzzle, we cannot overlook its potential impact on real scientific practice, both from a descriptive and a normative point of view. Considering the descriptive dimension, meta-scientific assumptions may influence how scientists assess certain modeling scenarios. For example, they can constrain how to evaluate the epistemic validity of models that do not maintain any connection with reality (i.e., holistically distorted models) or address the existence of a plurality of conflicting models. A researcher with strong realist commitments will assess the latter situation —e.g., cancer (Soto & Sonnenschein, 2021; Weinberg, 2007)— quite differently from someone with anti-realist, or pragmatist commitments. Here, the realism-driven researcher might be prompted to promote a unified investigation aimed at finding the alleged missing common cause. Considering the normative dimension, they may influence the formulation of criteria through which to assess the validity of the representations. A researcher with realist and veritistic commitments will require a connection with the world; another who refuses to embrace substantial metaphysical and epistemological commitments —i.e., a pragmatic standpoint— will simply resort to community and research needs. In short, the meta-scientific commitments adopted in answering SR-P, while theoretical in nature, may eventually impact how scientists conduct their research.

The theoretical and practical importance of meta-scientific assumptions calls for an assessment of their role within the debate on scientific representation.[1] I will contend that a specific meta-scientific platform has dominated the philosophical debate. Namely, onto-representationalism. This framework has been structured around three principles: representationalism, realism, and veritism. I will illustrate how a substantial number of accounts have implicitly adopted it in answering SR-P. Representationalism constitutes one of the ideas usually adopted by default. Roughly speaking, it holds that the epistemic status of models derives from their ability to faithfully represent, directly through shared features

[1] Hereafter I will discuss only the proposals of the advocates of the indirect view of representation (Levy, 2015; Toon, 2012). Notice that SR-P refers to the attempt to clarify the relationship between a model and a target system. By stating that models function as a prop in the game of make-believe which prescribes imaginings about T, advocates of direct representation simply do not address SR-P.

or indirectly through robust counterfactual patterns, certain aspects of the causal/mechanistic structure of phenomena (Knuuttila, 2011, p. 263). It is the representational relationship that allows us to explain the virtues and epistemic value of our best models. In turn, representationalism is based on two additional meta-scientific ideas: one ontological, realism, and the other epistemological, veritism. Our scientific models are representations of the target systems because they provide true information —veritism— and do so by accurately reflecting certain parts of the ontic structure of the phenomena in certain degrees and respects —realism. The combination of these three ideas has shaped the main meta-scientific platform adopted to offer a solution to SR-P.

My hypothesis is that onto-representationalism leads to an inadequate conceptualization of SR-P that negatively impacts scientific practice. More specifically, I will argue that the problem lies in the *semantic assumption,* which is usually overlooked. The latter is the one that allows us to determine which models are genuinely representative by providing a way to discriminate whether the linguistic items that structure the model system represent the non-linguistic items of the corresponding target system. I will show that this is an unavoidable problem consubstantial to this framework since it is impossible to renounce the semantic commitment, adopting only the representational, ontological, and epistemological ones. In other words, it is impossible to dissociate the assumptions that integrate onto-representationalism: the viability of one rests on the validity of the others. Therefore, if one wants to retain the representationalist, realist or veritist commitments, one must face the challenges raised by the semantic assumption: the articulation of a criterion that allows one to answer the question of what makes a representation accurate and, hence, true. This forces onto-representationalist to face the challenges inherent to the interpretation of language, a problem that the history of philosophical thought has shown to be extremely elusive.

To substantiate these analyses, I will discuss the different variants framed within onto-representationalism: classical and sophisticated. I will show that both suffer from deficiencies. The former, which grounds the representational relation on certain shared characteristics, faces the problem of accounting for the epistemic role and the representational character of models based on holistic distortions. The latter, who argues that our models can represent their corresponding target systems without literally representing the causal/mechanistic structure of the phenomenon, leaves the semantic problem unanalyzed, thus calling into question the validity of SR-P analysis.

In a nutshell, the aim of the paper is to analytically structure the theoretical assumptions underlying one of the meta-scientific frameworks commonly used to account for SR-P —i.e., onto-representationalism—, as well as to elucidate its main shortcomings, which I will locate in the semantic assumption. The paper will be structured as follows. In the second and third sections, I will articulate the fundamental theses that structure the onto-representational approach. In the fourth section, I will discuss the main arguments presented by the advocates of classical onto-representationalism. In the fifth section, I will introduce the attempts made by sophisticated onto-representationalism to overcome the limitations of the classical approach.

2. The Onto-representational Approach

There is great controversy about the nature of the representational relationship connecting model systems and their corresponding target systems. Anjan Chakravartty (2010) has made an analytical distinction between those authors who advocate an informational account and those who defend a functional account. The former (Bueno & French, 2011; Craver, 2014; Craver & Kaplan, 2020; da Costa & French, 2003; Giere, 1988; Strevens, 2008; van Fraassen, 1980), exponents of the so-called dyadic view, state that the representational relation is based on an objective parameter that is independent of any pragmatic consideration of the agents. This objective correspondence —usually articulated in terms of similarity or some kind of -morphism— allows scientists to obtain true information about the phenomenon under study. The latter (Frigg & Nguyen, 2020; Giere, 2006; Suárez, 2004; van Fraassen, 2008), advocates of the triadic view, deny that the representational relation can be conceptualized independently of the pragmatic interests and needs

of the cognitive agents. No representational relation exists until an agent takes certain correspondences to satisfy specific cognitive needs. To my mind, the situation is somewhat more complicated than that proposed by Chakravartty. I suggest the following analytical classification.

(1) A first group would include those authors who adopt an onto-representational approach to account for the epistemic value of models. Two variants can be distinguished: "classical onto-representationalism" (e.g., Giere, 1988, 2006; van Fraassen, 2008; Weisberg, 2007) and "sophisticated onto-representationalism" (e.g., Bokulich, 2018; Kuorikoski & Ylikoski, 2015; Rice, 2021; Verreault-Julien, 2021). The difference lies in the parameter proposed for establishing the representational relationship between the model and the target system. The former have argued that there must be certain shared characteristics, whether material or formal. For this reason, they have resorted to the idea of -morphism or similarity. The latter have contended that the existence of certain kinds of models —i.e., holistically distorted models— prevents the establishment of a direct representational relationship. Consequently, they have suggested an alternative representational parameter: universal patterns of behavior based on stable relations of counterfactual (in)dependence that take place between certain variables in the system. What is important is that both approaches are based on a common meta-scientific platform: onto-representationalism. Note that there is no incompatibility in introducing pragmatic aspects and defending an onto-representational perspective: the agent decides what to represent and how; the world determines the epistemic virtues of representations.

(2) The second group would be integrated by those authors who refuse to adopt onto-representationalism to account for SR-P (e.g., Elgin, 2017; Knuuttila, 2021). These authors have challenged some assumption, whether realism, representationalism, or veritism. This is the reason why they conceptualize the representational relation differently. For example, Elgin speaks of exemplification. Knuutila frames the problem in artifactual terms; depending on the media and modes of presentation, the representational relation will be one or another. To evaluate the adequacy of the models, these authors offer an epistemic standard related to the social practices of justification.

(3) A third group would be composed of those authors who avoid making substantive commitments of a meta-scientific nature, just emphasizing the pragmatic component of representation (e.g., Frigg & Nguyen, 2020; Suárez, 2004). Obviously, this is not a unitary stance. For example, Suárez has argued that no basic parameter connects the model systems with their corresponding target systems in a general way (Suárez, 2003). Inference, he says, constitutes the superficial characteristic of scientific representation. Thus, a model represents its target if users are able to draw inferences from the phenomenon through the model. On the other hand, Contessa has deemed it necessary to explain why a user can employ a model to make valid inferences about a target system (Contessa, 2007, p. 57; see also Contessa, 2011, pp. 126-27). Thus, he has offered a substantive proposal of representation that places the notion of interpretation as the necessary and sufficient condition of epistemic representation. What binds these different accounts is their belief that an adequate response to SR-P does not involve any substantive commitment to principles of a philosophical nature —e.g., realism or veritism.

(4) A fourth group would gather those authors who argue that it is not necessary to talk about representation in order to render our modeling practices meaningful (e.g., de Oliveira, 2022). Models are not meaningful because they represent reality, but because their use allows us to learn about the world. Through a material engagement with them, cognitive agents are able to scaffold certain types of activities to solve research problems, making sense of reality. The problem of scientific representation consists in analyzing the interactions that exist between modelers and models, not the representational relationship.

Here, I will focus on onto-representational proposals (1) since they have dominated and shaped much of the debate. Advocates of onto-representationalism *sensu lato* believe that one of the most valuable goals of scientific practice is to produce faithful representations of reality that allow us to advance our understanding of the world. To achieve them, it is necessary to obtain true information about the aspect of interest of the phenomenon we are modeling. Let's say, for example, to understand the role of a particular cyclin-dependent protein kinase in the cell cycle, the dynamics of a population, the behavior of a gas, or the evolution of an economic system. This usually requires the construction of model

systems: more manageable devices —whether abstract or material— that help researchers reason about the target system by reducing its complexity. The manipulation and study of these systems allow scientists to obtain true information about certain aspects considered critical to the behavior of the target system.

Representationalists consider that merely possessing bits of true information about the aspect or behavior of interest is insufficient. Modelers must assemble the pieces of information into explanatory narratives that successfully answer specific why questions —usually codified as research goals. This calls for providing explanations that detail the causes, mechanisms, or patterns of behavior that make sense of the emergence of the (aspect/behavior of the) phenomenon considered relevant within the research domain (Craver, 2014; Glennan, 2017; Machamer *et al.*, 2000; Strevens, 2008). The representational function is linked to the explanatory function. A model is representational iff it includes an explanation that accounts for a certain aspect of the world. Note that the representation of that aspect of the world can be done either directly or indirectly. The former strategy will consist in providing an explanation describing the causes or mechanisms involved in the production of the phenomenon —classical onto-representationalism. The latter will aim to provide an explanation that uncovers universal patterns of behavior. That is, invariance relationships that hold between variables —sophisticated onto-representationalism. Consider the following example. Cancer cell invasion is typically assumed to be a protease-dependent process: cancer cells secrete proteases that degrade the basement membrane (BM) and invade the extracellular matrix. In recent years, 3D cell culture models have suggested the possibility of protease-independent invasion (Glentis *et al.*, 2017). This statement derived from the model will be true iff, indeed, it is effectively confirmed in real systems, and not only in model systems or the results prove to be replicable in different kinds of cell cultures. Note that the appeal to truth is essential: only models that provide true information/explanations are epistemically adequate. For the onto-representationalist, truth constitutes the *standard of epistemic acceptability*. It provides the criteria to judge whether the descriptions comprising the model system, or the inferences drawn from it, are adequate and valuable. As can be appreciated, advocates of onto-representationalism suggest an extremely tight connection between representation, explanation, and truth.

I think it is possible to trace the origins of this explanationism, and its close link with onto-representationalism, to the Ontic Conception of explanation, originally put forward by José Alberto Coffa in the late 1970s and extensively developed by Wesley Salmon (Bokulich, 2018; Salmon, 1984; Wright, 2015, 2018). It contends that the world possesses a causal structure independent of the mind. Explanations, rather than constituting arguments, or representations, are objective entities that subsist *de re*, being independent of any pragmatic consideration (Craver, 2014, p. 40; Strevens, 2017). This view has been opposed by proponents of the so-called epistemic conception, who argue that explanations should be thought of in essentially representational terms (e.g., Bechtel & Abrahamsen, 2005). Recent work on scientific modeling has forced advocates of the ontic conception to make certain concessions to the epistemic view. Two reasons can be advanced. First, it seems clear that explanations, rather than being objective elements directly perceived or grasped, are the epistemic result of the manipulation of models constructed by cognitive agents. Second, it is necessary to have clear criteria to distinguish good explanations from bad ones (Craver & Kaplan, 2020; Kaplan & Craver, 2011). Alisa Bokulich has argued that these considerations led proponents of the ontic conception to focus on *how explanations work* rather than what they are, thus emphasizing their agential and representational character (2016, p. 263). Once this semantic turn takes place, the "ontic" —the objective portions of the structure of the world— is no longer perceived as elements that can be accessed directly. They are now grasped indirectly through *veridical representations*. What Bokulich has not noted is that this has blurred the boundaries that initially separated the ontic conception from the epistemic one, giving rise to an approach that fuses intuitions belonging to both approaches: onto-representationalism. From this perspective, adequate models would be those that veridically represent certain aspects of the ontic structure of the phenomenon.

As could be appreciated, realism constitutes one of the central pillars of the onto-representational approach (Bokulich, 2016; Giere, 2008; Rice, 2021; Strevens, 2008). Although realism is part of onto-representationalism, there are crucial differences that separate both positions. The most central one is that realism does not necessarily commit itself to epistemological and semantic theses. One could adopt a minimal form of realism by accepting only the ontological one.

Take, for example, Devitt (1991), Asay (2018), or Leeds (2007). On the contrary, onto-representationalism demands the unconditional adoption of these three assumptions. In fact, the validity of one depends on the validity of the others. Representationalism needs veritism because it provides a way to discriminate what representations are more adequate: those that provide explanations whose propositional elements faithfully represent certain aspects of the world. Veritism provides a standard for discriminating between good and bad representations; truth. In turn, veritism needs representationalism because having accurate representations is how we get closer to a truer picture of reality. Veritism needs realism because only by postulating the existence of an already structured mind-independent reality is it possible to objectively differentiate good representations from bad ones. Realism requires veritism because otherwise, we would lack a way of determining whether representations genuinely refer to the world.

There is still a loose end. The advocate of onto-representationalism must still explain how we can determine whether, in fact, our representations accurately reflect aspects of reality. In other words, how the standard of evaluation delineated by veritism is implemented. One cannot appeal to pragmatic criteria, e.g., our explanations and representations are good because they satisfy our cognitive goals or social standards. This would lead to a sort of instrumentalist, pragmatist, or artifactual perspective that would violate the realist and veritistic foundations. It is necessary to make a semantic ascent by placing the idea of truth at the center of the analysis. Otherwise, veritism would be incomplete, and realism would be completely emptied of content since we would have no way of differentiating which representations reflect the ontic structure of phenomena and which do not. In the next section, I will reconstruct how onto-representationalism solves this problem.

3. The Semantic Dimension of the Onto-representational Approach

To make sense of the "representational character" of the so-called "onto-representationalist" approach we need to draw on a semantic thesis. The latter is required in order to materialize a standard of epistemic acceptability through which to differentiate operationally between true and false representations. Only then, veritist, realist, and representationalist assumptions will become intelligible. As I will show in this section, the onto-representationalist is compelled to resort to a very particular and problematic semantic theory of truth: the correspondence theory of truth.

Arguably, one might say that it is more convenient to appeal to epistemic considerations in order to avoid conceptual conundrums associated with truth. However, the latter play no substantial role in assessing the truth of a representation; they deal with justification practices taking place within a community of inquiry. Naturally, such a criterion of acceptability is unsuitable for the onto-representationalist: it makes it impossible to account for the representational character by sidelining the realist and veritist assumptions. Consequently, the representational status of scientific products can only be accounted for insofar as one shows that they stand for the things in the world to which they refer. In contrast to justification, truth is an essentially semantic concept: it points to an objective relation between certain linguistic items and an extralinguistic reality. This is the reason why the onto-representationalist needs to appeal to a semantic theory that explains how the idea of truth can have an operative materialization.

So far, the only theory that has offered a systematic explanation of how the propositions that structure the explanations that confer representativeness to scientific products can correspond to the parts of reality to which they refer, discriminating the correct representations from the incorrect ones, is the correspondence theory of truth (Rasmussen, 2014). The essential idea is that truth involves a relationship between two qualitatively different entities, a truthbearer and a truthmaker[2] (Engel, 2002; Goldman, 1999; Vision, 2004). A proposition (truthbearer) is true iff represents the corresponding state of the world (truthmaker) relevant to the content of the proposition (Marino, 2006). It should be noted

[2] There is a heated debate on what should be the truthbearers (ideas, beliefs, utterances, sentences, mental representations) and the truthmakers (world, facts, states of facts, tropes).

that truth does not refer to the content or meaning of a proposition. It is a property that depends on the relation that a proposition maintains with a certain element or phenomenon of the world (Ingthorsson, 2019). Meaning provides the conditions the proposition must meet to be true —truth-criteria. If these conditions are fulfilled, then it is possible to claim that the proposition is true (Burgess & Burgess, 2014). This demands that an adequate relation exists between the meaning of a proposition and the corresponding state of the world to which it refers. If this is the case, we can conclude that the propositions that structure the explanations derived from the model are true and, therefore, representative.

Resorting to a correspondence theory logically follows from the ontological, epistemological and representational commitments assumed by the onto-representationalist. After all, the latter is the attempt to formulate, in analytical and strictly formal terms, the fundamental intuition underlying onto-representationalism. Namely, there exists a differentiation between true and false representations, the former being those that hold an adequate relation with an objective non-linguistic reality independent of the mind. Christoph Kelp has summarized this idea as follows:

> Second, phenomena come with structure [...]. After all, for every phenomenon, no matter what its metaphysical nature might be, there is a set of true propositions that describe it. Structures help regiment the set of true propositions describing a phenomenon. It is true propositions about the structural relations between its elements and true propositions about intrinsic properties of the structure of the phenomenon that matter. (Kelp, 2021, p. 101)

Explanations that confer the representational status may be partially erroneous (false items or inadequate relations are included), incomplete (items of knowledge or relations between them are missing) or idealized (only certain explanatory roles are taken into account, leaving others aside). In the first two cases, the epistemic value of this kind of explanations would simply consist in pointing out what kind of work should be done to achieve a more accurate understanding of the mechanism or causal pattern under consideration. Authors such as Craver have already addressed this issue by introducing the idea of a "mechanism sketch". In the third case, the requirement of precise representation is still maintained for the difference maker analyzed. *Accurate representation* is necessary because truth, by definition, is an absolute fact that does not admit degrees: "That truth is absolute —there is, strictly, no such thing as a proposition's being more or less true; propositions are completely true if true at all. (Absoluteness)" (Wright, 1998, p. 60). It would be odd to say that a model represents approximately a certain property, aspect, or behavior. Someone might argue, "well, statements like 'my model roughly represents the molecular mechanism by which the Warburg effect occurs' make sense and are perfectly adequate". However, if we pressed our interlocutor a bit harder and forced him to make explicit the various statements that structure "the molecular mechanism underlying the Warburg effect" we would see that, in fact, the first statement is no longer tenable. Either the model represents how a particular oncogene is involved in such a mechanism or we do not know it. Either we know the particular effects of this oncogene downstream in the signaling network or we do not.

Onto-representationalism cannot circumvent truth by appealing to the concept of accuracy. The strategy would consist in maintaining that the concept of accuracy is much more general than that of truth. Unlike truth, a representation can be more or less accurate. We can speak of a graduation in accuracy, something that does not apply to the concept of truth. A result can be more or less accurate. On the other hand, a statement cannot be more or less true. Either it is true, or it is not. Does this mean it is feasible to adopt accuracy as a criterion of correctness instead of truth? This does not seem to be the case. When we claim that a given representation accurately describes a characteristic, causal pattern, or mechanism, we are specifying under what conditions the model is true. In other words, we are detailing the representational content of the model, which is composed of a series of statements or propositions about the aspect of the world to which it is addressed (Fish, 2021, p. 38). The central point is that the statements or propositions that structure the representational content cannot be judged as more or less true. If its truth conditions are satisfied, that is, if the aspect of the world to which it refers is how it is told to be, then it will be true. Otherwise, it will be false. The accuracy of the representation will ultimately depend on the correctness of the representational content, which will depend on the truth or falsity of the propositions that structure that content. In this case, accuracy leads us to the truth.

Similarly, it is not possible to resort to a deflationary concept of truth. As Kitcher pointed out (2012), if we assume that truth constitutes the standard to judge, objectively, the correctness of our representations and that it is reality, and not some sort of intersubjective agreement, that determines the epistemic validity of the propositions contained in our models —by virtue of being in an adequate and accurate representational relation—, then it is impossible to dissolve the problem of correspondence by appealing to some sort of deflationary theory of truth such as minimalism, redundancy, or primitivism. Otherwise, the onto-representationalist would be unable to explain the following point: What makes it possible that our models are successful and provide true information about the world or phenomena? It is not enough to show that the models are, in fact, true and provide knowledge; a reason must be given that explains why or in what sense this is so. In other words, it is necessary to specify the nature of the "representation-reality" relation. One cannot simply affirm that such a relation between propositions and the world does, in fact, take place; it is necessary to explain it. Otherwise, the knowledge or understanding of reality derived from models would be fortuitous and contingent since we would not have an objective criterion to distinguish the correct representations from the incorrect ones, thus being able to justify the epistemic value of the former. This would completely render realism and veritism meaningless.

For the sake of conceptual clarity, I find it convenient to make explicit the contributions of the semantic assumption to the development of the onto-representational approach, as well as its connections with representationalism, veritism, and realism. So, let me analytically reconstruct the theses of onto-representationalism:

> *1. Representationalist Assumption.* There is a class of models that are more valuable: representational models. Onto-representationalist considers that a model system represents its corresponding target system iff the former holds a certain representational relationship with the latter. This relation is articulated through a particular representational parameter, such as similarity, -morphism or universal patterns, at least concerning the causally central elements that explain the emergence or stability of certain aspects or behaviors considered relevant within a given domain of investigation.
> *2. Epistemological Assumption (veritism).* To be genuinely representational, the descriptions of the considered aspects of the target system must be articulated as explanations. To be sound —true—, the propositional elements that structure such explanations must accurately represent a certain class of mechanisms/processes or behaviors of the ontic structure of the phenomenon, those that account for the emergence of the aspect considered relevant —difference-maker. Therefore, there is an objective criterion to discriminate between correct and defective models.
>
> *3. Ontological Assumption (realism).* There is an objective and structured reality whose existence is independent of the mind. The phenomena —target systems— of the world that scientists study possess an already defined ontic structure independent of any pragmatic considerations. The latter guarantees the conditions of possibility for the existence of a representational connection between our models and the world. Moreover, it is the one that sets the necessary conditions for scientists to be able to configure an objective criterion through which to differentiate, at the epistemological level, true explanations from false ones.

As de Oliveira (2021, 2022) rightly points out, ontological and epistemological assumptions go hand in hand. Only if phenomena have an objective ontic structure that can be grasped through models is it possible to obtain knowledge of the former by studying and making use of the latter (2021, p. 2010). However, de Oliveira has overlooked the fact that these two assumptions are insufficient to articulate an adequate account that explains the representationalist character. Recall that one of the distinctive features of onto-representationalism is the indissoluble link between the representational, ontological, epistemological, and semantic theses. I have argued that the onto-representationalist must specify how we can assess whether the elements of the model systems stand in the right relation to those of the target system. Only in this way is it possible to justify why they are epistemically valuable, thus differentiating which models are genuinely representational. Onto-representationalists cannot simply remain on an epistemic plane. If so, the status of this meta-scientific stance would be in question. A semantic ascent is necessary. Correspondence theory enters into the pic-

ture to guarantee its suitability and adequacy: it allows the concrete materialization of the epistemological assumption by clarifying the role of truth in judging the correctness of representations.

> *4. Semantic Assumption (correspondence theory).* The epistemic virtues of the representations are explained because the propositional elements that structure the explanations—truthbearers—refer to the mechanisms/processes or behaviors that articulate the ontic structure of the target systems—truthmakers—, at least the difference-makers. To justify the correctness of this correspondence relation between linguistic and non-linguistic items, it is not enough to appeal to a deflationary theory of truth; it is necessary to appeal to some kind of correspondence theory. This is possible because of the prior commitment to the existence of an already structured objective reality.

On the basis of this analytical reconstruction, to offer a satisfactory account of scientific representation the onto-representationalist should provide an answer to each of the questions that structure the following scheme:

> Scientific Representation-Problem Scheme (SR-P Scheme): What is a scientific representation? It consists of:
>
> 1. *Coordination Problem (CP):* What is the representational parameter by which a model system successfully represents its corresponding target system?
>
> Classical Onto-representationalism: similarity or -morphism. Model systems represent their corresponding target systems iff there is a direct correspondence between certain components or characteristics, at least regarding the elements that are causally central to the emergence of the phenomenon under study (difference-makers).
> Sophisticated Onto-representationalism: counterfactual inferences. Model systems represent their corresponding target systems iff they uncover universal patterns of behavior. That is, if they allow us to obtain information about the counterfactual (in)dependence relationships that take place between certain characteristics/parameters that are considered relevant in the production of the *explanandum*.
>
> 2. *Accuracy Problem (AP):* What makes a representation accurate? Here is where "the problem" of the "scientific representation problem" really lies. Keep in mind that accuracy leads to truth.
>
> Classical Onto-representationalism: correspondence theory.
> Sophisticated Onto-representationalism: correspondence theory.

This scheme aims to synthesize the steps that must be taken in order to account for the cognitive role of representations. Note that I do not intend to claim that all proposals must answer to this scheme. Only those framed within onto-representationalism. In fact, authors such as Chakravartty (2010), Suárez (2004), Contessa (2007) or Frigg and Nguyen (2020) have considered that both questions are logically independent. For them, a model X can be inaccurate while being a representation of the target system T. Similarly, for advocates of a non-onto-representational approach it would not make sense to speak of accuracy, but rather of justification.

4. Classical Onto-representationalism

In the following two sections, I will show how the versions of onto-representationalism address SR-P. Here, I will delve into the response provided by classical onto-representationalism to CP and AP, exposing its problems.

Two parameters are invoked to elucidate the connection between the model system and the target system: similarity (e.g., Craver & Kaplan, 2020; Giere, 2006; Strevens, 2008; Weisberg, 2013) or morphism (e.g., Bueno & French, 2011; da Costa & French, 2003). The reason for proposing these parameters is to circumvent any linguistic consideration, thus avoiding the correspondence theory. As I will show, this strategy ultimately fails.

Advocates of similarity acknowledge that the connection between idealized model systems and target systems is complex and indirect: they can enter into many types of relationships. What makes it possible to connect the two systems is specification and instantiation. Specification refers to how the model is described (through linguistic, diagrammatic, mathematical, or computational devices, among others). In other words, it is the step that connects the model and the description of the model, conferring the former its representational status (Godfrey-Smith, 2006, p. 733). The description of the model also needs to be instantiated. Instantiation involves determining on which aspects of the world the model system focuses (conventions are sometimes needed because models do not naturally map to parts of real phenomena), which elements of the model represent which elements of the world, and what standards of accuracy should be used to evaluate the model. Representative model systems are similar to their corresponding target systems regarding certain characteristics to some degrees considered relevant by the community. Note that the similarities established between the model and the target are ambiguous until they are made explicit. This is where hypotheses come into the picture. Giere argued that hypotheses are statements that establish how a fully interpreted and specified model fits a particular real system (Giere, 1988, pp. 81-82). In other words, they are linguistic devices that highlight the relevant similarities, specifying to what degree they satisfy the specified requirements. Otherwise, similarity relations would be epistemically ambiguous. Because of onto-representational commitments —realism and veritism— the evaluation of these hypotheses requires a semantic ascent. The correct representation will be the one that, regardless of the pragmatic context of evaluation, stands in the proper relation to (certain aspects of) the phenomenon of the world. Consequently, even if the similarity relations are not articulated in terms of truth or falsity, since these entities (model and target system) are essentially non-linguistic, the same is not true when it comes to the hypotheses. Because of the realist and veritistic commitment, hypotheses must be evaluated by appealing to a semantic theory of truth.

Despite attempts to dissociate from any appeal to considerations of a linguistic nature, classical onto-representationalism is eventually forced to appeal to a correspondence theory to offer a substantive answer to the problem of scientific representation. Consequently, they must confront internal problems associated with correspondence theory. But not only that, they must also face another problem: the widespread use of idealizing assumptions and distortions. This issue can be easily overcome. Michael Strevens (2008, 2013, 2017) has argued that, while the elements of the model that detail the causal structure of the phenomenon must be veridical to be genuinely explanatory and representative, not all of them have to be so: only those that make the difference to the emergence of the phenomenon. Strevens seems to offer a partial answer to AP: models are genuinely representative iff they accurately reflect, at least, the elements that make a difference to the emergence of the phenomenon —difference-makers. There are, however, several problems with this solution.

As Rice (2019) has pointed out, this suggestion is based on a highly problematic assumption called "decomposition strategy". Here, it is taken for granted that: it is possible to decompose the phenomenon by isolating the contributions of those features that are central to the occurrence of the phenomenon (Target Decomposition Assumption); it is possible to decompose the model by differentiating the contributions of the precise parts from the idealized ones (Model Decomposition Assumption); and it is possible to connect the former with the latter (Mapping Assumption). However, this analytical decomposition is usually not so easy to achieve. Even when possible, one would still face the problem of holistically distorted models. Holistically distorted models are characterized by two features. First, they bear no representational relation to their respective target systems. Even the elements considered causally central to the production of the phenomenon are idealized (Rice, 2021). Second, this distortion is necessary: it allows the implementation of mathematical modeling tools to obtain information that would otherwise be impossible. This makes it infeasible to decompose the model by differentiating the contributions of the precise parts from the idealized ones, connecting the former with characteristics, causal patterns, or mechanisms in the world.

Extremely relevant proposals have been articulated around the notion of morphism. Morphism is essentially a relation between two structures. The main drawback of this approach is that, as Frigg (2023) points out, target systems, or phenomena, are not mathematical structures. To claim that a set-theoretic structure is morphic to a part of the physical world is a category mistake. What theoretical models represent are data models, which can indeed be treated as set-the-

oretic structures. However, to assume that our best models represent only data models, but not the corresponding phenomena, would completely empty the onto-representational perspective of its content (van Fraassen, 2008, p. 258). But not only that, theoretical models are abstract structures of set theory that lack representational content by themselves. The agent decides which structure to represent when specifying the system. Therefore, in order to acquire a representational dimension, structure-generating descriptions and interpretations are indispensable (Morrison, 2007, p. 207). These considerations bring us back to the initial problem; the interpretation of language. In addition, the advocate of morphism should also address the problem of holistically distorted models.

5. Sophisticated Onto-representationalism

Acknowledging the enormous difficulties faced by classical onto-representationalism, several authors have considered that only by analyzing the SR-P problem from an alternative conceptual standpoint is it possible to offer an adequate answer to AP while maintaining the onto-representational assumptions. In other words, they have tried to accommodate the existence of holistically distorted models through a renewed representational parameter that avoids linguistic considerations. I will refer to this conceptual position as "sophisticated onto-representationalism".

What differentiates the sophisticated onto-representationalist approaches from the classical ones is their interest in separating two seemingly indissoluble theoretical assumptions: representationalism and literalism. While the classical approach holds that it is possible to literally represent certain aspects of the ontic structure of phenomena, the sophisticated one considers that the access to this ontic structure is indirect: our best models represent reality, but not literally. This approach has been motivated by the widespread use of holistically distorted models. In sum, the goal of sophisticated onto-representationalist has consisted in making sense of the existence of holistically distorted models, avoiding any invocation of a correspondence theory while preserving onto-representational commitments.

The immediate question that arises is the following: if to be qualified as representational is no longer required to literally reflect aspects of the target system, then what is the parameter of representation? In other words, what is the answer to CP? Certainly, appeals to elements such as isomorphism or similarity *must* be avoided, since these require that the model system and the target system have some *shared features*. The sophisticated onto-representationalist holds that the representational relation must be characterized in *counterfactual inferential terms*. More specifically, in counterfactual (in)dependency relations. Modal inferential reasoning becomes the representational parameter to establish the model system-target system relationship. At first glance, it might seem that, by avoiding the appeal to shared characteristics, the advocates of sophisticated onto-representationalism are in a more favorable position to deal satisfactorily with the AP problem, since they would evade correspondence theory while making sense of the holistically distorted models. To analyze whether sophisticated onto-representationalism succeeds in overcoming the problems of its classical counterpart, I will consider the proposals of Bokulich and Rice.

Bokulich has articulated an onto-representational account called "Eikonic Conception": "I grant that explanation and understanding are "success terms", in that they require getting something right about the way the world is, and more generally, I take the eikonic conception of explanation to be compatible with a broadly realist approach to science" (2018, p. 796; see also 2016, p. 261). For Bokulich, scientists never study the pure phenomenon in its complexity. When investigating a phenomenon, scientists make use of a series of conceptual, methodological, or theoretical tools framed within a particular research program to generate representations of the target system. In other words, scientists do not study the phenomenon-in-the-world but a particular conceptualization dependent on the context and research interests; a phenomenon-as-represented. The explanations derived will depend on the latter. Let me analyze what answer Bokulich gives to the SR-P scheme:

> *CP*. Models need not literally reflect the elements of the target system to be considered genuinely representational. That is, they need not share characteristics with the target system to be representative. For Bokulich, maximizing realism (in the sense of literalism) in many cases neither leads to an increase in prediction accuracy nor to an improvement in the explanation that can be obtained. So, what are the parameters that establish the representational character? Basically two. First, the interests and epistemic resources of the scientists. Whether a model is representational depends on how the phenomenon is conceptualized and the properties scientists want to explain. These considerations derive from the distinction established by Bokulich between phenomena-in-the-world and phenomena-as-represented. Recall that it is not incompatible to offer an onto-representational account with introducing the pragmatic interests of the agents. Second, and more importantly, the model must provide scientists with modal information about patterns of counterfactual (in)dependence on the *explanandum* they aim to explain: answering how particular changes in the properties of the model would lead to specific outcomes. If the model allows us to correctly answer questions "What-if-things-would-have-been-different", then it is genuinely explanatory —not simply an *ad hoc* phenomenological model— and provides a *factive* understanding of the phenomenon[3] (2016, p. 271). This conceptual shift allows Bokulich to accommodate the use of holistic distortions and fictional models while retaining onto-representationalist intuitions. However, one question remains, how do we evaluate the degree of accuracy in capturing counterfactual dependency patterns?
>
> *AP*. Bokulich states that in order to represent a phenomenon P, the counterfactual structure of the model M must be *isomorphic*, in the aspects considered relevant within the domain of investigation, to the counterfactual structure of P (Bokulich 2011, p. 43). That is, the model system must provide precise modal information on how the target system would behave if certain elements/variables of the target system were altered. This can be done only if it *reproduces* the counterfactual characteristics considered relevant to the target system within a particular domain of investigation. This answer seems to take us back to the problems that classical onto-representationalism was prey to.

Another author who has articulated a sophisticated approach is Rice. His account, labeled Realism of Understanding (2021), holds that science aims to achieve a *factive* understanding of the world. To genuinely understand a phenomenon, what is stated should be true, thus standing for the world. Like Bokulich, he argues that it is possible to achieve factive understanding through models that do not literally represent their corresponding target systems. That is, through holistically distorted models. Let us analyze Rice's response to SR-P:

> *CP*. While it is true that holistically distorted models are unable to represent in a literal way the characteristics, causes, or mechanisms that make the difference, these models have the potential to be considered genuinely explanatory and representational. The holistic distortions in these models allow modelers to use mathematical modeling tools to access information that would otherwise be impossible to have. What kind of information? Information about the patterns of counterfactual (in)dependence. These relationships of counterfactual (in)dependence are intended to uncover patterns of universality. Rice argues that universality means: "*the stability of certain patterns or behaviors* across systems that are heterogeneous in their features" (2021, p. 155). He introduces the concept of universality classes to highlight that model systems that are heterogeneous in their characteristics can exhibit the same patterns of behavior. Thus, Rice provides us with a global representational parameter: a model system represents its corresponding target system iff it provides true modal information about how certain changes in the system's characteristics/parameters alter (or not) the behavior of the phenomenon of interest. That is, if it provides information about counterfactual (in)dependence relationships backed by patterns of universality.

[3] Factive understanding accounts are committed to theses of a veritistic and realist nature (see, e.g., Baumberger *et al.*, 2017).

AP. Rice, unlike Bokulich, does not offer a clear answer to the question of what it means to provide adequate information about counterfactual (in)dependence relationships. It is only possible to claim that a model represents its corresponding target system and provides factive understanding if there is some way to properly connect the behavior of the ideal case (of the system described by the holistically distorted models) with the behavior of real phenomena. Rice only states:

> The key thing to notice is that holistically distorted models can provide accurate modal information because universality guarantees that the *model system's patterns of behavior will be similar to those of the target systems, even if the actual entities, causal interactions, and processes of those systems are extremely different*. Therefore, even if the model drastically and pervasively distorts the fundamental nature of the relevant features of real-world systems in order to use various mathematical modeling techniques, it can still be used to explain because *many of the patterns of counterfactual dependence and independence that hold in the model system will be similar to those of real-world systems.* (Rice, 2021, p. 161 my emphasis; see also p. 156)

Rice, therefore, does not clearly state how it is possible to assess whether the counterfactual (in)dependence patterns are, in fact, true and provide a factive understanding, i.e., whether or not they truly reflect those counterfactual (in)dependence relationships. In other words, he does not make explicit what is the standard of accuracy that allows us to determine whether our model is genuinely representative. It might be possible to deduce the answer from his writings: there must be some correspondence between the patterns of counterfactual (in)dependence of the model and those of the world. This would mean that Rice's view is still tied to the requirement of accurate representation, slipping back into the problems associated with the interpretation of language (see Carrillo & Knuuttila, 2022). From these analyses, one can conclude the following. Either Rice's proposal is incomplete and, therefore, he has to specify how it is possible to answer the questions raised above without committing, like Bokulich, the same mistakes of classical representationalism, or simply his proposal does not offer a substantial improvement over classical representationalism since it also leads to a correspondence theory.

In sum, the sophisticated onto-representationalist seems ultimately to fail in the attempt to dissociate "representationalism" from "literalism", thus being able to provide an adequate response to SR-P. I believe that the inability of the onto-representationalist to provide a satisfactory answer to SR-P is not explained by the argumentative weakness of the concrete proposals that comprise it. Rather, it is the product of the assumptions that integrate it: representationalism, realism, and veritism. The tight linkage of these assumptions leads irremediably to a semantic question difficult to solve: How are the elements or behaviors uncovered by the model system connected to those of the target system? The answer to this question is crucial since it provides a standard to differentiate the correct representations from the defective ones and, by extension, the models that are genuinely representational from those that are not. Ultimately, it is necessary to appeal to a semantic theory, embodied in some kind of correspondence theory. This is what makes it possible to relate the linguistic elements (the inferences/explanations associated with the model) with the non-linguistic elements (the mechanism/process or behavior of the modeled phenomenon).

The failure of onto-representationalism should not, on the other hand, lead us to a kind of skepticism about the possibility of achieving a satisfactory answer to SR-P. Delineating a concept of representation that allows us to discriminate between adequate and inadequate models is a desirable theoretical goal that can help us to articulate a more efficient scientific practice. Rather, the criticisms raised against this framework should lead us to ask ourselves whether the meta-scientific approach through which SR-P has usually been approached is adequate. At present, a plurality of authors are immersed in developing alternative meta-scientific platforms to onto-representationalism (see, e.g., de Oliveira, 2022; Knuuttila, 2011, 2021). These proposals simply do not assume the ontological and epistemological assumptions of onto-representationalism. Undoubtedly, these alternative frameworks will not be without problems and criticisms that

need to be critically addressed. However, the moral that I would like to draw from this analysis is that instead of trying to articulate ever more sophisticated and complicated onto-representational proposals, perhaps we should explore the possibilities offered by these new approaches regarding SR-P.

6. Conclusion

In this paper, I have analytically structured the representational, ontological, epistemological, and semantic assumptions underlying the meta-scientific platform that the majority of accounts that have attempted to address SR-P have adopted. Namely onto-representationalism. I have shown to what extent this meta-scientific framework leads to an inadequate conceptualization of SR-P. Drawing on current literature on scientific representation, I have argued that its main weakness, and the reason why it is not suitable for addressing SR-P, lies in the semantic assumption. I have contended that it is impossible to dissociate the four assumptions that integrate onto-representationalism. But not only that, I have shown that the validity of one depends on the validity of the others. If one assumption proves to be erroneous, the contribution of the others will be invalidated. Therefore, if one wants to retain the representationalist, realist, or veritistic commitments, one must face the challenges raised by the semantic assumption. That is, the articulation of a criterion that allows one to answer the question of what makes a representation accurate. This implies facing the challenges inherent to the interpretation of language, a task in which epistemologists have been immersed for years, and which correspondence theorists do not seem to have solved so far. This leads onto-representationalist approach to a stalemate: either they face the semantic problem, or they leave aside their representational, ontological, and epistemological commitments, thus exploring new avenues. Since the former option has turned out to be a dead-end so far, nothing prevents us from exploring alternatives to onto-representationalism. I have concluded that perhaps the best way to deal efficiently with SR-P is to adopt one of the alternative approaches to onto-representationalism that are currently being developed.

Acknowledgments

This work has been funded by the Consejería de Educación de la Junta de Castilla y León (Spain) and the European Social Fund. I thank Ana Cuevas Badallo, Obdulia Torres, Andrei Moldovan, Benedicto Acosta, Cristian Saborido and the reviewers for their comments.

REFERENCES

Asay, J. (2018). Realism and theories of truth. In J. Saatsi (Ed.), *The Routledge handbook of scientific realism* (pp. 383-393). Routledge.

Baumberger, C., Beisbart, C. & Brun, G. (2017). What is Understanding? An Overview of Recent Debates in Epistemology and Philosophy of Science. In S. Grimm, C. Baumberger & S. Ammon (Eds.), *Explaining understanding. New perspectives from epistemology and philosophy of science* (pp. 1-34). Routledge.

Bechtel, W. & Abrahamsen, A. (2005). Explanation: A Mechanistic Alternative. *Studies in the History and Philosophy of Biology and Biomedical Sciences*, 36(2), 421-441. https://doi.org/10.1016/j.shpsc.2005.03.010

Bokulich, A. (2011). How scientific models can explain. *Synthese,* 180, 33-45. http://dx.doi.org/10.1007/s11229-009-9565-1

Bokulich, A. (2016). Fiction As a Vehicle for Truth. *The Monist,* 3, 260-279. https://doi.org/10.1093/monist/onw004

Bokulich, A. (2018). Representing and Explaining: The Eikonik Conception. *Philosophy of Science*, 85(5), 793-805. https://doi.org/10.1086/699693

Bueno, O. & French, S. (2011). How theories represent. *British Journal for the Philosophy of Science,* 62(4), 857-894. https://doi.org/10.1093/bjps/axr010

Burgess, A. & Burgess, J. (2014). *Truth*. Princeton University Press.

Carrillo, N. & Knuuttila, T. (2022). Holistic idealization: An artifactual standpoint. *Studies in History and Philosophy of Science*, 91, 49-59.

Chakravartty, A. (2001). The Semantic or Model-Theoretic View of Theories and Scientific Realism. *Synthese*, 127, 325-345.

Chakravartty, A. (2007). *A metaphysics for scientific realism.* Cambridge University Press.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science,* 74(1), 48-68. https://doi.org/10.1086/519478

Contessa, G. (2011). Scientific Models and Representation. In S. French & J. Saatsi (Eds.), *The Continuum companion to the philosophy of science* (pp. 120-137). Continuum.

Craver, C. (2014). The Ontic Account of Scientific Explanation. In M. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences. The case of biology and history* (pp. 27-52). Springer.

Craver, C. & Kasplan, D. (2020). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. *The British Journal for the Philosophy of Science*, 71(1), 287-319. https://doi.org/10.1093/bjps/axy015

de Oliveira, G. S. (2021). Representationalism is a dead end. *Synthese,* 198, 209-235.

de Oliveira, G. S. (2022). Radical artifactualism. *European Journal for Philosophy of Science*, 12, 36.

Devitt, M. (1997). *Realism and truth.* Princeton University Press.

Elgin, C. Z. (2017). *True enough*. The MIT Press.

Engel, P. (2002). *Truth*. Routledge.

Fish, W. (2021). *Philosophy of perception. A contemporary introduction*. Routledge.

Frigg, R. (2023). *Models and theories. A philosophical inquiry*. Routledge.

Frigg, R. & Nguyen, J. (2020). *Modelling nature: An opinionated introduction to scientific representation.* Springer. https://doi.org/10.1007/978-3-030-45153-0

Giere, R. (1988). *Explaining science. A cognitive approach.* The University of Chicago Press.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science,* 71(5), 742-752. https://doi.org/10.1086/425063

Giere, R. (2006). *Scientific perspectivism.* The University of Chicago Press.

Glennan, S. (2017). *The new mechanical philosophy.* Oxford University Press.

Glentis, A., Oertle, P., Mariani, P., Chikina, A., El Marjou, F., Attieh, Y., Zaccarini, F., Lae, M., Loew, D., Dingli, F., Sirven, P., Schoumacher, M., Gurchenkov, B. G., Plodinec, M. & Vignjevic, D. M. (2017). Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. *Nature communications*, 8(1), 924. https://doi.org/10.1038/s41467-017-00985-8

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21, 725-740. http://dx.doi.org/10.1007/s10539-006-9054-6

Goldman, A. (1999). *Knowledge in a social world.* Oxford University Press.

Kaplan, D. & Craver, C. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science,* 78(4), 601-627. https://doi.org/10.1086/661755

Kelp, C. (2021). *Inquiry, knowledge, and understanding.* Oxford University Press.

Kitcher, P. (2012). On the explanatory role of correspondence truth. *Philosophy and Phenomenological Research,* 64(2), 346-364. https://doi.org/10.1111/j.1933-1592.2002.tb00005.x

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, 42(2), 262-271.

Knuuttila, T. (2021). Models, Fictions and Artifacts. In W. Gonzalez (Ed.), *Language and scientific research.* Palgrave Macmillan.

Kuorikoski, J. & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, 192(12), 3817-3837.

Leeds, S. (2007). Correspondence truth and scientific realism. *Synthese,* 159, 1-21. http://dx.doi.org/10.1007/s11229-006-9064-6

Levy, A. (2015). Modeling without models. *Philosophical Studies*, 152(3), 781-798.

Machamer, P., Darden, L. & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science,* 67, 1-25. https://doi.org/10.1086/392759

Marino, P. (2006). What should a correspondence theory be and do? *Philosophical Studies,* 127(3), 415-457.

Morrison, M. (2007). Where have all the theories gone? *Philosophy of Science,* 74, 195-228. https://doi.org/10.1086/520778

Peng, L., Trucu, D., Lin, P., Thompson, A. & Chaplain, M. A. (2017). A Multiscale Mathematical Model of Tumour Invasive Growth. *Bulletin of mathematical biology*, 79(3), 389-429. https://doi.org/10.1007/s11538-016-0237-2

Rasmussen, J. (2014). *Defending the correspondence theory of truth*. Cambridge University Press. https://doi.org/10.1017/CBO9781107415102

Rice, C. (2019). Models don't decompose that way: a holistic view of idealized models. *The British Journal for the Philosophy of Science*, 70(1), 179-208.

Rice, C. (2021). *Leveraging distortions. Explanation, idealization, and universality in science.* The MIT Press.

Salmon, W. (1984) *Scientific explanation and the causal structure of the world*. Princeton University Press.

Soto, A. M. & Sonnenschein, C. (2021). The cancer puzzle: Welcome to organicism. *Progress in Biophysics and Molecular Biology*, 165, pp. 114-119.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A,* 44(3), 510-515. http://dx.doi.org/10.1016/j.shpsa.2012.12.005

Strevens, M. (2017). How idealizations provide understanding. In S. R. Grimm, C. Baumberger & S. Ammon (Eds.), *Explaining understanding: New essays in epistemology and the philosophy of science* (pp. 37-49). Routledge.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, 71, 767-779. https://doi.org/10.1086/421415

Suárez, M. (2010). Scientific Representation. *Philosophy Compass,* 5(1), 91-101. https://doi.org/10.1111/j.1747-9991.2009.00261.x

Toon, A. (2012). *Models as make-believe: Imagination, fiction and scientific representation.* Palgrave Macmillan.

van Fraassen, B. (1980). *The scientific image.* Oxford University Press.

van Fraassen, B. (2008). *Scientific representation: paradoxes of perspective*. Clarendon Press.

Verreault-Julien, P. (2021). Factive inferentialism and the puzzle of model-based explanation. *Synthese,* 199, 10039-10057.

Vision, G. (2004). *Veritas. The correspondence theory and its critics*. The MIT Press.

Weinberg, R. A. (2007). *The biology of cancer*. Garland Science.

Weisberg, M. (2007). Who is a Modeler. *The British Journal for the Philosophy of Science*, 58, 207-233.

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world.* Oxford University Press.

Woodger, J. H. (1929). *Biological principles. A critical study*. Routledge.

Wright, C. (1998). Truth: A Traditional Debate Reviewed. *Canadian Journal of Philosophy Supplementary Volume*, 24, 31-74.

Wright, C. (2015). The Ontic Conception of Scientific Explanation. *Studies in History and Philosophy of Science Part* A, 54, 20-30. http://dx.doi.org/10.1016/j.shpsa.2015.06.001

Wright, C. (2018). Ontic Explanation Is Either Ontic or Explanatory, but not both. *Ergo: An open Access Journal of Philosophy,* 5, 997-1029. http://dx.doi.org/10.3998/ergo.12405314.0005.038

Mariano Martín Villuendas is a research fellow at the Department of Philosophy, Logic, and Aesthetics – Instituto Universitario de Estudios de la Ciencia y la Tecnología at the University of Salamanca. His main areas of research involve issues related to the epistemology and history of the Life Sciences.

Address: Department of Philosophy, Logic, and Aesthetics – Instituto Universitario de Estudios de la Ciencia y la Tecnología, University of Salamanca. Edificio I+D+I, Calle Espejo n.º 2, 37007, Salamanca, Spain. Email: marianomv@usal.es – ORCID: https://orcid.org/0000-0002-6814-7346

Theoria

An International Journal for Theory, History and Foundations of Science

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Summary

Volume 40/1, January 2025, pp. 1-104

Articles

Matías Gariazzo (Universidad de la República), «Against Pritchard's refutation of epistemic relativism», *Theoria*, 2025, 40/1, 68-80

This paper reconstructs Duncan Pritchard's (2011; 2021; 2025) refutation of epistemic relativism and presents an objection to it. This refutation presupposes that epistemic relativism would be true if there were rationally irresolvable deep disagreements. Our objection shows that the examples of deep disagreement Pritchard presents have features that, while making them rationally resolvable, not all deep disagreement has.

Mariano Martín-Villuendas (Mariano Martín-Villuendas), «The unfeasibility of onto-representationalism», *Theoria*, 2025, 40/1, 81-97

The aim of the paper is to analytically structure the framework commonly adopted to address the Scientific Representation Problem (SR-P), i.e., onto-representationalism, and to examine its main problems. Due to its very theoretical conditions, I conclude that onto-representationalism constitutes an inadequate meta-scientific platform to approach SR-P. I locate the problem in the semantic assumption.

Referees for Theoria
Informantes de Theoria
(2023-2024)

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Referees for THEORIA (2023-2024)
Informantes de THEORIA (2023-2024)

The editors gratefully acknowledge the contribution of the referees listed below. They all reviewed one or more manuscripts submitted between 2023 and 2024. Their reports helped the members of our Editorial Board in making their final decisions.

El equipo editor agradece la colaboración de las personas evaluadoras que se detallan debajo. Cada una de ellas emitió informes sobre uno o varios manuscritos recibidos en el periodo 2023-2024. Sus informes sirvieron para fundamentar las decisiones del Consejo editorial sobre su publicación.

Martín Abreu Zavaleta
Manuel Almagro
Ramón Alvarado
Miguel Álvarez Lisboa
Gloria Andrada de Gregorio
Mattia Andreoletti
Brandon James Ashby
Gunnar Babcock
Sergio Balari
Lorenzo Baravalle
Áxel Barceló
Eduardo Barrio
Lucien Baumgartner
Claus Beisbart
Francesca Bellazzi
Karl Bergman
Lilian Bermejo
Reinaldo Bernal
Joan Bertrán San Millán
Brandon Boesch
Florian Boge
Paolo Bonardi
Marcelino Botín
Mark Bowker
Vojislav Božičković
Adam Bradley
Liam Bright
María Caamaño
Fabrizio Calzavarini
Fabrizio Calzavarini
Alejandro Cassini
Dan Cavedon-Taylor
María Cerezo
Hasok Chang
Tony Cheng
Pablo Cobreros
Dimitri Coelho Mollo
Cristina Corredor
Ana Cuevas Badallo
André Curtis-Trudel
Silvia De Toffoli
Fabio Del Prete
Hugh Desmond
Dennis Dieks
José Antonio Díez Calzada
Chao Ding
Louis Doulas
Zoe Drayson
Laura Duhau
Santiago Echeverri
Matthias Egg
Gretchen Ellefson
Marco Facchin
Marco Facchin
Boaz Faraday Schuman
Giulia Felappi
Gabriele Ferretti
Aldo Filomeno
Sarah Fisher
Samuel C. Fletcher
Carolina Flores
María José Frapolli
Doreen Fraser
Antonio Gaitán
Enrico Galli
Manuel García Carpintero
María José García-Encinas
Pío García
José Ángel Gascón
Santiago Ginnobili
David Glick
Bernardo Gonçalves
Vincent Grandjean
Alex Gregory
Daniel Gregory
Bojana Grujičić
Francesco Guala
Charles H. Pence
Nina Haket
Peter Hanks
Margherita Harris
Joyce Havstad
Manuel Heras Escribano
Tyler Hildebrand

Inês Hipólito
Federico Holik
Robert Howell
Fabian Hundertmark
Josh Hunt
Manuel J. García-Pérez
Victor J. Luque
Christian Jendreiko
Ethan Jerzak
Juan Pablo Jorge
Yannic Kappes
Eddy Kemming Cheng
Alex Kerr
David Kinney
Kepa Korta
Decio Krause
Victoria Lavorerio
Dustin Lazarovici
Olivier Lemeire
Juliana Lima
Michelle Liu
Olimpia Lombardi
Guy Longworth
Cristian López
Fabrizio Macagno
Lilith Mace
Josep Macià
Paolo Maffezioli
Pablo Magaña
P.D. Magnus
Stephen Mann
María Laura Martínez
María Martínez Ordaz
Fernando Martínez-Manrique
Carlo Martini
Claudio Masolo
Lucy Mason
Dana Matthiessen
Kael McCormack
José Jorge Mendoza
Tushar Menon
Marcin Miłkowski
Martin Montminy
Alice Murphy
Jesús Navarro Reyes
Ilkka Niiniluoto
Elias Okon
Andrea Onofri
Ismael Ordóñez Miguéns
Ezequiel Paez
Michele Palmira
Graham Parsons
Ross Patrizio
Andrew Peet
Diana Pérez
José Antonio Pérez Escobar
Saúl Pérez González
Eduardo Pérez-Navarro
Bryan Pickel
Vicente Pico
David Pineda
Sophie R. Allen
Nicholas Rimell
J. Brendan Ritchie
Ariel Roffé
Aida Roige
Luis Rosa
Stéphanie Ruphy
Quentin Ruyant
Pablo Rychter
Alexander Sager
Catharine Saint-Croix
Léa Salje
Gui Sanches de Oliveira
Silvia Carolina Scotto
Miguel Ángel Sebastián
Johanna Seibt
Markus Seidel
Giuseppe Sergioli
Yafeng Shan
Ashley Shaw
Joshua Shepherd
Kenneth Silver
Léna Soler
Aris Spanos
Katie Steele
Javier Suárez
Pilar Terrés
Justin Tiehen
Robert Trueman
Matthew Tugby
Joaquín Miguel Valdivielso Navarro
Íñigo Valero
Víctor M. Verdejo
Agustín Vicente
Massimiliano Vignolo
Cristina Villegas
Bryan W Roberts
Sam Wilkinson
Malte Willer
Benjamin Young
Nick Young
José Zalabardo
Henrik Zinkernagel

GUIDELINES FOR AUTHORS

THEORIA will only consider original papers in the areas of competence of our editorial board, never exceeding 10.000 words. We will try to make a decision on your paper within eight weeks.

We will only accepts electronic submissions. For further Instructions, please, visit http://www.ehu.es/theoria

EDITORIAL PROCEDURE

THEORIA is a peer-reviewed journal. The editors, together with the competent members of the editorial board, will screen the submissions and decide whether to send them to referee. Every paper will be then assessed by two (blind) reviewers. On this basis, the editors will make the final decision. The list of reviewers is regularly published in the journal. Our archives in the editorial platform are open to research agencies (such as the Spanish FECYT) for external verification of the quality of the review process.

COPYRIGHT NOTICE

THEORIA is published under a Creative Commons Licence. Authors who publish with this journal agree to the following terms:

- Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.
- Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal.
- Authors are permitted and encouraged to post their work online (e.g., in institutional repositories or on their website) prior to and during the submission process, as it can lead to productive exchanges, as well as earlier and greater citation of published work.

SUBSCRIPTION AND BACK ISSUES

THEORIA is published 3 times a year. Annual subscription rates for 2010 are: €50 for individuals and €70 for institutions. Single Issues can be ordered for €20 each. A Cumulative Index covering the numbers 21-50 is available for €3. For ordering information or any enquiry concerning subscriptions, please contact: publicaciones@ehu.eus.

Publishers

This journal is published jointly by the University of the Basque Country and the Center for Analysis, Logic, and Legal Informatics (CALIJ) in Donostia/San Sebastián, Spain.

CALIJ

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

THEORIA is associated with
the Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España (SLMFCE)
and the Sociedad Española de Filosofía Analítica (SEFA)

THEORIA

An International Journal for Theory, History and Foundations of Science

Vol. 40/1 • January 2025

Dep. Legal: SS - 820/85 ISSN 0495-4548 e-ISSN 2171-679X