HUNGARIAN PHILOSOPHICAL REVIEW

VOL. 67. (2023/2)

The Journal of the Philosophical Committee
of the Hungarian Academy of Sciences

Teleology: Old Wine in New Skins

Edited by László Bernáth – Dániel Kodaj

Contents

Teleology: Old Wine in New Skins

Teleology is rarely discussed in contemporary philosophy outside of a few highly specialized areas. One of those is the history of Aristotelian thought, where teleology is seen as a central feature of the biological and human world, indeed, of substances in general. Another is the philosophy of biology, where experts debate the viability and proper shape of a modern, naturalized conception of teleology, one that is compatible with Darwinism but still allows us to talk about functions and goals with respect to organisms. Both of those research programs can suggest that robust teleology has no theoretical relevance any more (it belongs in the philosophical museum) and that goal-directedness, to the extent that it exists, is best analyzed in a reductive naturalist framework.

The present collection tries to buck this trend by offering alternative views about the history, meaning, and contemporary relevance of teleology. The collection begins with *Darwin and Design* by Michael Ruse, one of the founders of modern philosophy of biology, who unravels the complicated history of real and imaginary tensions between Darwinism and theistic/organicist theories of nature. Gergely Kertész's *On the Status of Teleological Discourse* carries the topic forward to the present day and argues that teleology can be seen as a real and important phenomenon even in a mainstream naturalist framework. Erik Åkerlund's *Models of Finality* takes us back to pre-Darwinian times into the thick of Aristotelian natural philosophy and it outlines three distinct models of final causation (exemplified by Aristotle, Buridan, and Averroes, respectively), demonstrating that 'the' pre-Darwinian teleological view of nature is not a monolith but a multifaceted philosophical movement. Capitalizing on insights from that movement, Gyula Klima argues in "Teleology, Intentionality, Naturalism" that a Thomistic conception of voluntary action continues to be much more persuasive than its modern physicalist counterparts.

The second half of the collection makes systematic use of teleology in the context of contemporary metaphysics, normative ethics, and action theory. In "The Metaphysics of Spooky Teleology", Daniel Kodaj seeks to construct a definition of robust teleology in the context of analytic metaphysics. Mohsen Moghri's paper, *An Axiological Ultimate Explanation for Existence*, discusses contemporary theories of cosmic teleology as a response to the question of why there is something rather than nothing. From metaphysics and cosmology, we move to normative ethics in *The Aporia of Categorical Obligations and an Augustinian Teleological Way Out of It* by László Bernáth, who defends categorical ob-

ligations against attacks from modern moral philosophy and offers a conception of categorical obligations that is both historical and novel. Finally, in *Intentional Actions and Final Causes*, Ferenc Huoranszki argues that mainstream action theory cannot explain the difference between an agent's actions and those events that merely happen to her; in contrast, a view that involves intrinsic goal-directedness is ideally suited for that purpose.

The reader will also find two papers that do not belong to the thematic collection but have been in the pipeline for a while, waiting for the next English issue of the Hungarian Philosophical Review. Attila Hangai's *What is Rational Reconstruction in the History of Philosophy?* is a reflection on the 2022 English issue of HPR, which was devoted to the historiography of early modern philosophy. Ayumu Tamura, in *The Role of Experience in Descartes' Metaphysics*, examines Descartes' conception of experience, with special attention to the claim that Descartes identified experience with intuition and understanding.

We would like to thank the editorial board of the Hungarian Philosophical Review, especially Gergely Ambrus, for wholeheartedly supporting this project since its inception and for tolerating our inability to keep any deadlines. We also thank the John Templeton Foundation and the Ian Ramsey Centre for a grant that supported a two-year research program at CEU *(Meant to Be: Resuscitating the Metaphysics of Teleology)*, laying the groundwork for the present collection. Preparing this special issue was also supported by an OTKA (Hungarian Scientific Research Fund) grant awarded by the National Research Development and Innovation Office under its Postdoctoral Excellence Programme (PD131998), as well as by an OTKA research grant (K132911). Finally, we are extremely grateful to our anonymous reviewers for contributing to the natural goal of all academics, the publication of peer-reviewed papers.

László Bernáth – Dániel Kodaj

Michael Ruse

Darwin and Design*

> Let us recognize Darwin's great service to Natural Science in bringing back to it Teleology: so that, instead of Morphology *versus* Teleology, we shall have Morphology wedded to Teleology. (Gray 1874)

> What you say about teleology pleases me especially, & I do not think anyone else has ever noticed the point. I have always said you were the man to hit the nail on the head. (Darwin, letter to Asa Gray June 5, 1874, DCP-LETT-9483)

I. THE DESIGN ARGUMENT

The Argument from Design, or the Teleological Argument, is one of the oldest and best-known – often taken to be the most compelling – arguments for the existence of God (Ruse 2017). Not just God, but a God of Christianity, who is All-Powerful, All-Knowing, and All-Loving. It is to be found in Plato's *Phaedo*, the dialogue supposedly reporting on Socrates' last day on Earth. "One day I heard someone reading, as he said, from a book of Anaxagoras, and saying that it is Mind that directs and is the cause of everything. I was delighted with this cause, and it seemed to me to be good, in a way, that Mind should be the cause of all. I thought that if this were so, the directing Mind would direct everything and arrange each thing in the way that was best" (Cooper 1997. 97 c-d). So, now one has a guide to understanding and, as a bonus, a guide to discovery. "Then if one wished to know the cause of each thing, why it comes to be or perishes or exists, one had to find what the best way was for it to be, or to be acted upon,

* I want to acknowledge the two incredibly helpful (anonymous) readers of an earlier version of this paper. Thanks to them, the paper is much improved. In the course of a very long academic career (60 years), I have been touched again and again by the generosity my fellow philosophers have shown towards me, the care and attention put into their comments. As even I come to the end of a career, it is people like these who have made my life such a joy and meaningful. Thank you, all.

or to act." Aristotle, Plato's successor, did not have anything akin to the Christian God. His ultimate cause, the Unmoved Mover, spent its time doing the only thing open to a truly perfect being, contemplating its own perfection! It had no knowledge of the physical world, including us (Sedley 2008). Aristotle, however, followed Plato in seeing our world as deeply purposeful – the hand exists to grasp things, the rain exists in order to fertilize the ground. Meaning by "ultimate" reason *why* something happened and by "proximate" reason *how* something happened, for Plato, if the ultimate reason for the purpose was the Form of the Good, the proximate reason for him – and others including Aristotle and then the Christian's – was that the world in some sense is an organism. Plato's *Timaeus* was on this very topic, with the Designer being the "Demiurge," aka the Form of the Good. First, that the Designer worked for the good. "Now surely it's clear to all that it was the eternal model he looked at, for, of all the things that have come to be, our universe is the most beautiful, and of causes the craftsman is the most excellent. This, then, is how it has come to be: it is a work of craft, modeled after that which is changeless and is grasped by a rational account, that is, by wisdom" (Cooper 1997; *Timaeus* 29a). Plato does not regard this creation – the universe – to be some dead, lifeless entity. It is a living being with a soul. "Now why did he who framed this whole universe of becoming frame it? Let us state the reason why: He was good, and one who is good can never become jealous of anything" (29d-e). Clearly the God being himself good had to model things on the best, the Form of the Good. And this brings in intelligence. And so straight off we get a world soul. "Guided by this reasoning, he put intelligence in soul, and soul in body, and so he constructed the universe. He wanted to produce a piece of work that would be as excellent and supreme as its nature would allow. This, then, in keeping with our likely account, is how we must say divine providence brought our world into being as a truly living thing, endowed with soul and intelligence" (30b-c). Aristotle likewise bought into this picture of the world as an organism. He distinguished proximate causes or "motor" causes, those that make things happen, from final causes, the reason for things to happen. (Better known is Aristotle's four-part division of causes: efficient, material, formal, and final. However, when dealing with organisms, he brews this down to a two-part division: proximate and final (Aristotle 1984 a, b).) In the case of the organism, for instance, the proximate cause is the rain bringing the seed to life. The final cause, the reason for the proximate cause, is the flowering plant attracting insects to fertilize it. Not having a designer, or Designer, in the sense of Plato, Aristotle inclined rather to see the whole world as alive, in some sense, within itself. Hence, there is a kind of vital force directing things towards perfection, that is the Unknown Mover (which in some sense is a perfect being). In more recent times, people spoke of an entelechy or *élan vital*.

Of course, living four hundred years before Jesus, neither Plato nor Aristotle were Christians. But Plato's "Mind" or God was the Form of the Good, the

source of all knowledge and that which is of value. Christians, particularly the greatest theologian of all, St Augustine, identified this Form with their God, noting that as for the Christian God, the Form of the Good was not merely all powerful and knowing, as well as all good, but outside the physical world – eternal and never changing. Note that the organism is not to be identified with the Creator/Designer. That would be unacceptable pantheism. The organism is the result of the efforts of the Creator/Designer, as in Genesis One. ("In the beginning God created the heaven and the earth" 1:1). With this organic metaphor as background, the Christians took up the argument from design with fervor. St Thomas Aquinas gave the classic exposition. Note that, although he was much influenced by Aristotle's thinking on final causes – bodies "act for an end" – ultimately, he, as a Christian, is forced back to a kind of Platonic Great Designer in the Sky.

> The fifth way is taken from the governance of the world. We see that things which lack intelligence, such as natural bodies, act for an end, and this is evident from their acting always, or nearly always, in the same way, so as to obtain the best result. Hence it is plain that not fortuitously, but designedly, do they achieve their end. Now whatever lacks intelligence cannot move towards an end, unless it be directed by some being endowed with knowledge and intelligence; as the arrow is shot to its mark by the archer. Therefore some intelligent being exists by whom all natural things are directed to their end; and this being we call God. (Aquinas 1981, *Summa Theologiae* Ia.q2.a3)

Generations of undergraduates, who have read Aquinas only in extracts such as this, come away with the belief that this is the end of things. Not true! As a Christian, Aquinas always thought faith took precedence over reason, as used in the Fifth Way. Jesus made that very clear. Remember the encounter with the disciple Thomas who doubted that Jesus had been resurrected.

> Then he said to Thomas, "Put your finger here; see my hands. Reach out your hand and put it into my side. Stop doubting and believe."
> Thomas said to him, "My Lord and my God!"
> Then Jesus told him, "Because you have seen me, you have believed; blessed are those who have not seen and yet have believed." (John 20: 27-29)

Aquinas pointed out that, without the supremacy of faith, the lazy and the ignorant would never get to know God (Ruse 2019). But the overall tenor was certainly that reason and evidence are high on the list of things acceptable to God and that, therefore, the organicist approach to understanding, of the world and of God, was very well taken.

II. CHANGING ROOT METAPHORS

What changed this? The three Rs! *Renaissance, Reformation, Revolution*. The *Renaissance*, going back to the wisdom of the Ancients, soon showed that not everyone was enamored by design. The Roman poet Lucretius, putting into verse older beliefs of the atomists and others, gave a vivid alternative picture.

> At that time the earth tried to create many monsters
> with weird appearance and anatomy –
> androgynous, of neither one sex nor the other but somewhere in between;
> some footless, or handless;
> many even without mouths, or without eyes and blind;
> some with their limbs stuck together all along their body,
> and thus disabled from doing harm or obtaining anything they needed.
> These and other monsters the earth created.
> But to no avail, since nature prohibited their development.
> They were unable to reach the goal of their maturity,
> to find sustenance or to copulate. (Sedley 2007. 150–153; *De rerum natura V 837-848)*

At first, nothing works, it is all a dysfunctional mess. Then, given infinite time, there is functional success.

> First, the fierce and savage lion species
> has been protected by its courage, foxes by cunning, deer by speed of flight. But as for
> the light-sleeping minds of dogs, with their faithful heart,
> and every kind born of the seed of beasts of burden,
> and along with them the wool-bearing flocks and the horned tribes,
> they have all been entrusted to the care of the human race… (V 862-867)

No design. Just chance and lots of time. Even if this seems implausible at first, it lodges in the mind and is worrisome.

The *Reformation*, with its emphasis on *sola scriptura*, obviously downplayed reason in favor of faith. Luther even went so far as to refer to reason as a "whore"! There were some responses. Some passages of the Bible seem best interpreted in terms of design. There was King David's contribution, the opening of Psalm 19: "The heavens declare the glory of God; and the firmament sheweth his handiwork." Saint Paul also rushed briefly over the idea: "For the invisible things of him from the creation of the world are clearly seen, being understood by the things that are made, even his eternal power and Godhead; so that they are without excuse" (Romans 1:20). But this is indeed slim pickens given the overall length and scope of Holy Scripture. Another, more sociological response, was that of the English. The second half of the sixteenth century saw the long

reign of Elizabeth the First, and – much desired after the short reign of Bloody Mary, who tried to enforce Catholicism on her reluctant subjects – the consolidation of Britain as a Protestant nation. England's initial break from Rome was done more for political than theological reasons. Henry wanted to divorce his Catholic wife so he could marry Anne Boleyn on the hope of getting a male heir. When the Pope refused, Henry picked up his country and went home – less metaphorically, took Britain out of the Catholic realm and into the Protestant. Truly, then, *sola scriptura* never had the hold on the English that it had on the Protestant countries of Europe. (Scotland also, given the influence of the Calvin follower, John Knox.) Something theologically distinctive and convincing was needed for the English, and the gap was filled with a distinctively English form of natural theology, one that emphasized the analogy between nature and the many efficient machines that the English were now inventing and using (Ruse 2003).

Overall, however, notwithstanding the English, *sola scriptura* was a strong clarion call. And this fit nicely with the (Scientific) *Revolution*, usually dated from 1543 and the publication of Copernicus' heliocentric picture of the universe – *De Revolutionibus Orbium Coelestium*[1] – to 1687 and the publication of Newton's causal theory, *Philosophiae Naturalis Principia Mathematica*.[2] As historians stress, above all the revolution was one of change of metaphors, from the already-encountered "world as an organism," to the newcomer: "world as a machine."

> At all times there used to be a strong tendency among physicists, particularly in England, to form as concrete a picture as possible of the physical reality behind the phenomena, the not directly perceptible cause of that which can be perceived by the senses; they were always looking for hidden mechanisms, and in so doing supposed, without being concerned about this assumption, that these would be essentially the same kind as the simple instruments which men had used from time immemorial to relieve their work,… (Dijksterhuis 1961. 497)

Robert Boyle (1627–1691), physicist and philosopher, was explicit: the world is "like a rare clock, such as may be that at Strasbourg, where all things are so skillfully contrived that the engine being once set a-moving, all things proceed according to the artificer's first design, and the motions of the little statues that as such hours perform these or those motions do not require (like those of puppets) the peculiar interposing of the artificer or any intelligent agent employed by him, but perform their functions on particular occasions by virtue of the general and primitive contrivance of the whole engine" (Boyle 1686. 12-13). The world now was seen simply as a contraption, governed by eternal, unchanging

[1] https://en.wikipedia.org/wiki/De_revolutionibus_orbium_coelestium
[2] http://www.gutenberg.org/ebooks/28233

laws, simply going through the motions, without rhyme or reason. Of course, you might say that machines have purposes. A guillotine is hardly for slicing tomatoes. But within the context of science, this part of the metaphor was dropped. There were to be no ends, no final causes, things that the philosopher Francis Bacon likened to Vestal Virgins, beautiful but barren. And this means that the world is value free. It is just dead substance in motion, and any values we find are values we ascribe to it. The heart has no value as such, but value in the sense that we humans think it of value (because of its results). To the organicist, it is just silly to say the heart has no intrinsic value. Of course, it does – value to be found out there in the world. Value put there by a benevolent Creator (Plato), or part of the very fabric of the world (Aristotle). Since the root metaphor is the organism, the world is usually seen as developing, increasing in value. Few, if any organicists, would pull back from the inference that we humans are of the greatest value. The mechanist would undoubtedly agree with this conclusion; but, think the value we put on humans is the value we put on humans, not something we find ready-made (Ruse 2021).

III. THE PROBLEM OF ORGANISMS

Mechanism triumphant! There was however a rather large fly in the ointment. Organisms. The traditional argument from design covers both the organic and the inorganic. The hand exists in order to grasp; the rain exists in order to fertilize. But it had always been recognized that the appearance of design is far less in the inorganic than the organic. This said, Aristotle was not naive. He was fully aware that it is at times proper to speak of things as being accidental or contingent. He didn't think that an eclipse of the moon is necessarily for any great purpose. Is this just an exception to final cause thinking? Not really. The eclipse as eclipse is not a substance. Heavenly beings move in circles because that is the perfect figure and so that is part of their nature. But the effects are not substances and so not necessarily explicable in terms of final cause. "Nor does matter belong to those things which exist by nature but are not substances; their substratum is the substance. E.g. what is the cause of eclipse? What is its matter? There is none; the moon is that which suffers eclipse. What is the moving cause which extinguished the light? The earth. The final cause perhaps does not exist" (Barnes 1984. 1649; *Metaphysics*, 1044b8–b12). Whatever. No one felt much worry about dropping final cause talk about the inorganic world. Organisms were different. They apparently continued to demand final-cause talk. The eye really is for seeing! The eye exists in order to see. The final cause of the eye is sight.

Faced with this problem, Robert Boyle played the philosophical equivalent of the three-card trick. He distinguished between acknowledging the use of final causes qua science and the inference qua theology from final causes to a

designing god. First: "In the bodies of animals it is oftentimes allowable for a naturalist, from the manifest and apposite uses of the parts, to collect some of the particular ends, to which nature destinated them. And in some cases, we may, from the known natures, as well as from the structure, of the parts, ground probable conjectures (both affirmative and negative) about the particular offices of the parts" (Boyle 1688. 18). Then, the science finished, one can switch to theology: "It is rational, from the manifest fitness of some things to cosmical or animal ends or uses, to infer, that they were framed or ordained in reference thereunto by an intelligent and designing agent" (Boyle 1688. 19). From a study in the realm of science, of what Boyle would call "contrivance," to an inference about design – or rather Design – in the realm of theology.

Organisms were booted out of science into the realm of religion. A solution, but hardly a satisfactory solution, for all that, over the next century or more, some good biological science was done thanks to this uneasy compromise. Naturalistic mechanistic thinking in the physical sciences. Religion-entwined organismic thinking in the biological sciences. As a result of this, the argument from design for the existence of God continued to flourish, particularly in Britain, dependent as its religion was on natural theology. (The state-sponsored Anglican religion. By the middle of the eighteenth century, more faith-centered religions were starting to appear in numbers. The Methodists particularly.) It is little surprise then that the classic exposition of the argument should appear at the beginning of the nineteenth century – Archdeacon Paley's *Natural Theology*.

> In crossing a heath, suppose I pitched my foot against a stone, and were asked how the stone came to be there; I might possibly answer, that, for any thing I knew to the contrary, it had lain there for ever: nor would it perhaps be very easy to show the absurdity of this answer. But suppose I had found a watch upon the ground, and it should be inquired how the watch happened to be in that place; I should hardly think of the answer which I had before given, that, for any thing I knew, the watch might have always been there. (Paley 1802. 1)

The watch shows organization, marks of design. The stone does not. Shall we simply say that the watch just happened? "Or shall it, instead of this, all at once turn us round to an opposite conclusion, viz. that no art or skill whatever has been concerned in the business, although all other evidences of art and skill remain as they were, and this last and supreme piece of art be now added to the rest? Can this be maintained without absurdity? Yet this is atheism" (13-14). Paley continues:

> This is atheism: for every indication of contrivance, every manifestation of design, which existed in the watch, exists in the works of nature; with the difference, on the side of nature, of being greater and more, and that in a degree which exceeds all

> computation. I mean that the contrivances of nature surpass the contrivances of art, in the complexity, subtility, and curiosity of the mechanism; and still more, if possible, do they go beyond them in number and variety; yet, in a multitude of cases, are not less evidently mechanical, not less evidently contrivances, not less evidently accommodated to their end, or suited to their office, than are the most perfect productions of human ingenuity.
>
> I know no better method of introducing so large a subject, than that of comparing a single thing with a single thing; an eye, for example, with a telescope. As far as the examination of the instrument goes, there is precisely the same proof that the eye was made for vision, as there is that the telescope was made for assisting it. They are made upon the same principles; both being adjusted to the laws by which the transmission and refraction of rays of light are regulated. (14-15)

The watch is designed. The eye is just like the watch. Hence the eye is designed. Or rather, Designed – by God!

IV. HUME AND KANT

There had been earlier criticisms of the argument, but ultimately these had not succeeded. Apparently devastating were some of the arguments of David Hume, in his *Dialogues Concerning Natural Religion*, published some twenty years earlier than Paley's *Natural Theology*. He showed that the traditional argument from design – the argument of Plato and Augustine and Aquinas – is riddled with problems. On the one hand, who is to say that there is only one designer, and who moreover is to say that this designer got things right straight off? Our experience of complex entities is that usually this is a group effort, drawing on the experience of many attempts, sometimes failures, sometimes successes, in the past. "But were this world ever so perfect a production, it must still remain uncertain, whether all the excellences of the work can justly be ascribed to the workman. If we survey a ship, what an exalted idea must we form of the ingenuity of the carpenter who framed so complicated, useful, and beautiful a machine? And what surprise must we feel, when we find him a stupid mechanic, who imitated others, and copied an art, which, through a long succession of ages, after multiplied trials, mistakes, corrections, deliberations, and controversies, had been gradually improving?" (Hume 1779. 77). And was it just one workman? "And what shadow of an argument . . . can you produce, from your hypothesis, to prove the unity of the Deity? A great number of men join in building a house or ship, in rearing a city, in framing a commonwealth; why may not several deities combine in contriving and framing a world?" The trouble is, of course, that you are reading in your conclusion – a unique, all-powerful deity – right into your premises and then thinking that you have discovered or proved something.

And yet, this said – and much more – in the end Hume equivocates. He may be a believer. And then again, he may not be.

> That the works of Nature bear a great analogy to the productions of art, is evident; and according to all the rules of good reasoning, we ought to infer, if we argue at all concerning them, that their causes have a proportional analogy. But as there are also considerable differences, we have reason to suppose a proportional difference in the causes; and in particular, ought to attribute a much higher degree of power and energy to the supreme cause, than any we have ever observed in mankind. Here then the existence of a DEITY is plainly ascertained by reason: and if we make it a question, whether, on account of these analogies, we can properly call him a mind or intelligence, notwithstanding the vast difference which may reasonably be supposed between him and human minds; what is this but a mere verbal controversy? (130)

General opinion, with which I concur, is that Hume is a classic case of someone caught on the problem of "inference to the best explanation." You have a number of options and you must choose the best. Usually, you do this by eliminating the least satisfactory, until you have only one left standing. Sherlock Holmes gives the classic statement. "When you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth." The trouble is that organisms do seem as if designed. It is impossible that they not be. So, improbable though it may be, there must be something to the God hypothesis. You must eliminate all those that make no reference to a Designer of some sort.

Immanuel Kant, in his third Critique, *The Critique of Judgement*, had a somewhat different take on things. As a good Newtonian, he was convinced that the world is ruled by unbroken law. The proper root metaphor for understanding is the machine metaphor. Yet, there is in organisms the undeniable appearance of design. And you cannot really do biology without this assumption of design. You would not be able to ask about the use of anything. Hence, uneasily, Kant concluded that thoughts of final cause had to be allowed, but they were purely heuristic and not part of the real science.

> The concept of a thing as in itself a natural end is therefore not a constitutive concept of the understanding or of reason, but it can still be a regulative concept for the reflecting power of judgment, for guiding research into objects of this kind and thinking over their highest ground in accordance with a remote analogy with our own causality in accordance with ends; not, of course, for the sake of knowledge of nature or of its original ground, but rather for the sake of the very same practical faculty of reason in us in analogy with which we consider the cause of that purposiveness. (Kant 1790. 247)

An answer, if not a terribly satisfactory answer. Perhaps out of frustration at the thin solution he offered, Kant showed that sometimes he was more human than

ethereal philosopher, by turning bitterly on the source of this frustration, biology. You want to make the life sciences equal to the physical sciences? Good luck! "[W]e can boldly say that it would be absurd for humans even to make such an attempt or to hope that there may yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention has ordered; rather, we must absolutely deny this insight to human beings" (271).

V. PROBLEMS WITH DESIGN

We enter the nineteenth century and turn towards Darwin and his *Origin of Species* (Ruse 1999). As we do so, it is well to remember an important point made by Thomas Kuhn in his *The Structure of Scientific Revolutions* (1962). Few, if any, accept Kuhn's extreme idealism, that when (what he calls) "paradigms" change, the world itself changes – the before and after paradigms are "incommensurable." To the contrary, as we shall see fully in the Darwinian case, there is clearly much continuity between before and after paradigms. However, Kuhn is clearly right that revolutions don't just happen. There must be reason for change and the most obvious reason is that the older paradigm is no longer functioning that well. It is coming apart with increasing visible problems and the virtue of the new paradigm is either that it can explain and hence eliminate the problems, or it can do an end run around the problems, so they are no longer so very pressing. We can think of the pre-Darwinian paradigm, not so much as "Creationism" in the sense of today's American biblical literalists – six-day creation, six thousand years ago, Adam and Eve in the Garden of Eden in their birthday suits – but Creationism in the sense of the design-like nature of the organic world precludes an explanation in terms of unbroken law. Miracles, divine interventions in the natural order of things, are needed to create already-functioning organisms. In the words of the polymath, historian and philosopher of science William Whewell:

> Geology and astronomy are, of themselves, incapable of giving us any distinct and satisfactory account of the origin of the universe, or of its parts. We need not wonder, then, at any particular instance of this incapacity; as for example, that of which we have been speaking, the impossibility of accounting by any natural means for the production of all the successive tribes of plants and animals which have peopled the world in the various stages of its progress, as geology teaches us. That they were, like our own animal and vegetable contemporaries, profoundly adapted to the condition in which they were placed, we have ample reason to believe; but when we inquire whence they came into this our world, geology is silent. The mystery of creation is not within the range of her legitimate territory; she says nothing, but she points upwards. (Whewell 1837/3. 587–588.)

But what if – quite independently of Darwin – the organic world is nothing like as design-like as these Creationists suppose? If someone, Charles Darwin, is going to offer an evolutionary account of the organic world, then the assumption is going to be that blind law can explain organisms in their entirety. If it cannot do this, because of the design-like nature of organisms, then evolution – the "Evolutionism" paradigm – is impossible. Obviously, at one level, the evolutionist like Darwin is going to have to explain that blind law can do the job. However, if there is no job to be done, then the evolutionist can win by default, as it were. No barriers.

As it happens, this fear of the Creationists is only too well placed. Even by the 1830s, people like Whewell were coming to realize that there are important aspects of organisms – not just ephemeral by-products – that seem to have no direct purpose (Whewell 1837; Ruse 1977). Aspects for which final-cause explanations simply seem neither needed nor appropriate. Most obvious were what, in the next decade, the anatomist Richard Owen (1849) was to call "homologies," the isomorphisms between organisms of very different species. The paradigm example is of the vertebrate forelimb. Very different organisms have the bone order and structure of their forelimbs – forelimbs where the uses are very different – in parallel. The arm of humans is used for grasping; the forelimb of the horse, for running; the wing of the bird for flying; the flipper of the porpoise for swimming; and more. There seems no purposeful reason for any of this.

This problem, as you might say, is internal to biology. Then for a worry more external to biology, by the 1850s, a decade before the *Origin* was published, Whewell started to fret about extraterrestrials. In an anonymously authored book, *The Plurality of Worlds* (1853), Whewell posed the question of whether we humans are unique. Or, if there are many planets through the universe that carry living beings, including living human-like beings? Why was Whewell worried about this? Quite simply because his revealed religion – the religion of faith and the Bible – was under threat from his natural religion – the religion of reason. The evidence of design, of which he made so much in his stand against evolution, works only if you see design out there. The less evidence of design, the less reason to invoke non-law bound causes. This rather suggests then that we should find purpose – final causes – everywhere. Meaning not only on our planet but throughout the universe. And the only point of other planets, the only possible purpose of them, is to support life. Hence, we expect to find life teeming everywhere. More than this, there is not much point in life if it does not lead, whether by evolutionary forces or otherwise, to intelligent beings of some form. But then comes the question of their relationship to the Creator. A multitude of Creators is hardly plausible. Unfortunately, if we do have intelligent beings elsewhere, this opens the possibility of their falling into sin as have we humans. Which means that God, presumably in the form of Jesus, has to come down to their planets in order to save them. We end with the theologically absurd

– absurd and obnoxious – conclusion that perhaps Jesus is being crucified on Friday– every Friday – somewhere in the universe, to save souls. An implication like this must be stopped, and the obvious way is to argue that, despite universal purpose, the existence of non-inhabited worlds, apparently pointless worlds, is nevertheless highly plausible.

In the course of his argument, Whewell brought several lines of fire to bear. Thus, he argued at some length that the geological record shows that, for much of the life of this Earth of ours, there was either no life or no intelligent life. Hence, concluded Whewell, there was no point to this world for much of its existence, at least, not in the sense of being designed for organisms in general and humans in particular. In a somewhat analogous manner, Whewell also pointed out that many aspects of organisms show no point, in the sense of being of any benefit to them. Thus, the nipples on the male are hardly of any value to anyone. Similarly, Whewell cited the homologous forms of the skeletons of man and sparrows, which hardly do anyone or anything very much good. And, in a passage anticipating Charles Darwin's discussion of the struggle for existence in the *Origin of Species*, Whewell drew attention to the fact that most organisms seem to have little point anyway, because they die before maturity: "to work in vain, in the sense of producing means of life which are not used, embryos which are never vivified, germs which are not developed, is so far from being contrary to the usual proceedings of nature, that it is an operation which is constantly going on, in every part of nature" (Whewell 1853. 248).

There were other arguments brought to bear on the case. God does not always work for direct organic benefit, but for other ends such as similarity, symmetry, and beauty. Hence, analogous structures (homologies) in different organisms exist "for the sake of similarity" (248). Similarly, the different hexagonal forms of snowflakes have no end but symmetry and beauty. And in addition to supplying different ends for God, Whewell made much play of a version of the Design Argument which he called the "Argument from Law." Even though we may see no direct ends, "the existence of Laws of Nature, governing and producing the phenomena of the universe, makes manifest to us the existence and operation of God" (251). Finally, in order to find some point to uninhabited other worlds, Whewell made a new suggestion – the most crucial of all for his revised position – namely that man's mind is in essential respects like God's Mind, and part of our task on Earth might be to bring ourselves closer to God by tracing His laws as manifested by the endless motions of the heavenly bodies.

> For if, on the earth, the Creator have placed a race who are not only endowed with a portion of the Divine Intellect, but who are placed there in order, (at least among other purposes,) that they may cultivate and develop this gift, and thus, rise nearer and nearer to the condition of the Divine Intellect, and be fitted, so far, for an immortal existence; we cannot have any ground to think that the scheme of creation

> is too narrow; or that it needs, in order to give it sufficient dignity and value, and a worthy object in our eyes, that other worlds should be stocked with races of creatures... (309)

As you might imagine, suggestions like this were embraced with all of the enthusiasm of facing a lead balloon. Sir David Brewster, Scottish physicist and biographer of Newton, countered with *More Worlds than One: The Creed of the Philosopher and the Hope of the Christian* (1854). He argued that there is intelligent life everywhere, including on the Sun! You can imagine how well this bolstered the case of the non-evolutionists.

VI. ON THE ORIGIN OF SPECIES

Turn now to Charles Darwin and his great work, *On the Origin of Species*, published in 1859. What did he try to do in that work? He tried to show that all organisms, living and dead, are descended from "one or a few forms," by a slow, natural – meaning law-bound – process that he called "natural selection." First, he talked about artificial selection, what the farmer practices on the stock and what fanciers do with their birds and dogs and other animals that they prize and want to improve. He shows that the secret is choosing and breeding from those that have the desired features, over and over, until those features are fixed in the line or group. He then moved to the natural world, arguing that natural populations always have lots of variation, a prerequisite for a selective process. Then come the two key chapters. First, the struggle for existence, showing that not all organisms can survive and reproduce.

> A struggle for existence inevitably follows from the high rate at which all organic beings tend to increase. Every being, which during its natural lifetime produces several eggs or seeds, must suffer destruction during some period of its life, and during some season or occasional year, otherwise, on the principle of geometrical increase, its numbers would quickly become so inordinately great that no country could support the product. Hence, as more individuals are produced than can possibly survive, there must in every case be a struggle for existence, either one individual with another of the same species, or with the individuals of distinct species, or with the physical conditions of life. It is the doctrine of Malthus applied with manifold force to the whole animal and vegetable kingdoms; for in this case there can be no artificial increase of food, and no prudential restraint from marriage. Although some species may be now increasing, more or less rapidly, in numbers, all cannot do so, for the world would not hold them. (Darwin 1859. 63–64.)

Then, in the next chapter, Natural Selection, he argued that the struggle within populations of organisms, with a range of variations, is going to lead to a natural selecting process.

> HOW will the struggle for existence, discussed too briefly in the last chapter, act in regard to variation? Can the principle of selection, which we have seen is so potent in the hands of man, apply in nature? I think we shall see that it can act most effectually. Let it be borne in mind in what an endless number of strange peculiarities our domestic productions, and, in a lesser degree, those under nature, vary; and how strong the hereditary tendency is. Under domestication, it may be truly said that the whole organisation becomes in some degree plastic. Let it be borne in mind how infinitely complex and close-fitting are the mutual relations of all organic beings to each other and to their physical conditions of life. Can it, then, be thought improbable, seeing that variations useful to man have undoubtedly occurred, that other variations useful in some way to each being in the great and complex battle of life, should sometimes occur in the course of thousands of generations? If such do occur, can we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection. (80–81)

The key point is that natural selection doesn't just lead to change. It leads to change in the direction of features that help their possessors. A faster lion after prey is going to do better than a slower lion. A darker moth on a sooty tree is better camouflaged than a lighter one. A hardier plant in a rough environment is going to do better than a more delicate one. Organisms will develop features, "adaptations", that help in the struggle for existence, or more importantly struggle for reproduction.

> How have all those exquisite adaptations of one part of the organisation to another part, and to the conditions of life, and of one distinct organic being to another being, been perfected? We see these beautiful co-adaptations most plainly in the woodpecker and missletoe; and only a little less plainly in the humblest parasite which clings to the hairs of a quadruped or feathers of a bird; in the structure of the beetle which dives through the water; in the plumed seed which is wafted by the gentlest breeze; in short, we see beautiful adaptations everywhere and in every part of the organic world. (60–61)

Darwin answered his question:

> I have called this principle, by which each slight variation, if useful, is preserved, by the term of Natural Selection, in order to mark its relation to man's power of selection. We have seen that man by selection can certainly produce great results, and can adapt organic beings to his own uses, through the accumulation of slight but useful variations, given to him by the hand of Nature. But Natural Selection… is a power incessantly ready for action, and is as immeasurably superior to man's feeble efforts, as the works of Nature are to those of Art. (61)

VII. IMPLICATIONS

Stop right here and make three important points. First, Darwin is offering a natural, law-bound, within-the-machine-metaphor explanation of those characteristics like the hand and the eye that supporters of the organic metaphor claim can be explained only within their perspective. This implies those that think natural, machine-like explanations of adaptations (the kind Kant ruled out as impossible) cannot be and one must rely on non-natural interventions, miracles, are wrong. Darwin says that the eye, for example, exists and works because those would-be sighted animals that had variations more efficient in the direction of sight survived and reproduced and those that did not, did not. Blind, unguided law all the way.

Second, as he and Asa Gray realized, Darwin was not eliminating teleological – final cause – explanations. He was giving an answer other than miracles, but he was giving an answer to the same problem – adaptive characteristics seem to refer to the future. However, for the Creationist, it was the Mind of God responsible – He saw the intended future and planned for it. For the Evolutionist like Darwin, it was a case of this worked in the past, let us assume it will go on working. Kant pointed out that we have a kind of repetitive cause and effect process. It is a matter of organization or even self-organization. "This principle, or its definition, states: An organized product of nature is that in which everything is an end and reciprocally a means as well. Nothing in it is in vain, purposeless, or to be ascribed to a blind mechanism of nature" (Kant 1790. 247–248). Darwin agrees, but he thinks that that is just the way things are. The eye leads to seeing leads to survival and reproduction leads to another eye and… the process keeps repeating, over and over again. Of course, we might be mistaken. Darkness might envelope the Earth and no one can see again; but, the Designer has the same problem.

The point is that there is a genuine reference to the future. Darwin is giving a naturalistic explanation of final cause. He is not denying it. Indeed, in the *Origin*, he uses the notion of final cause without need of qualification. He asks why cuckoos lay their eggs in the nests of others.

> It is now commonly admitted that the more immediate and *final cause* of the cuckoo's instinct is, that she lays her eggs, not daily, but at intervals of two or three days; so that, if she were to make her own nest and sit on her own eggs, those first laid would have to be left for some time unincubated, or there would be eggs and young birds of different ages in the same nest. (216–217, my italics.)

Continuing, supposing that this spaced-out laying would have disadvantages but that sometimes a cuckoo might lay its eggs in the nest of another bird:

> Now let us suppose that the ancient progenitor of our European cuckoo had the habits of the American cuckoo; but that occasionally she laid an egg in another bird's nest. If the old bird profited by this occasional habit, or if the young were made more vigorous by advantage having been taken of the mistaken maternal instinct of another bird, than by their own mother's care, encumbered as she can hardly fail to be by having eggs and young of different ages at the same time; then the old birds or the fostered young would gain an advantage. And analogy would lead me to believe, that the young thus reared would be apt to follow by inheritance the occasional and aberrant habit of their mother, and in their turn would be apt to lay their eggs in other birds' nests, and thus be successful in rearing their young. By a continued process of this nature, I believe that the strange instinct of our cuckoo could be, and has been, generated.

The crucial point, however, is that, whereas Whewell appeals to divine intervention – "says nothing but points upwards" – Darwin offers a naturalistic law-bound explanation. Natural selection!

The third point is that, without effort or the need of ad hoc explanations, Darwin can answer those problems about seeming exceptions to the design-like nature of organisms, most especially homologies. They are a function of common ancestry. Evolution does not start each generation afresh. It very often modifies what it has according to new needs. There are good reasons to go fast? Then take the horse option? Out of the jungle and onto the plains. You need to be able to look around you for predators and prey. Go the bipedal option, opening up your forelimbs for new, or much improved, functions. The important point is that whereas Whewell is constantly playing catch up – God worked through laws to exercise our minds sort of thing – the evolutionist, the Darwinian evolutionist in particular, has a ready explanation at hand. This is all very much in the tradition of Kuhn's analysis of scientific revolutions. The old paradigm gets into trouble – constantly coming up with ad hoc solutions to solve problems. The new paradigm deals with these problems briskly – they break down under the new modes of explanation – and the scientist can and does move on.

Move on, but note that the Darwinian agrees with the Creationist that it is function and final cause that comes first. Homologies and the like are explicable, but they are side effects.

> It is generally acknowledged that all organic beings have been formed on two great laws – Unity of Type, and the Conditions of Existence. By unity of type is meant that fundamental agreement in structure, which we see in organic beings of the same class, and which is quite independent of their habits of life. On my theory, unity of type is explained by unity of descent. The expression of conditions of existence, so often insisted on by the illustrious Cuvier, is fully embraced by the principle of natural selection. For natural selection acts by either now adapting the varying parts of each being to its organic and inorganic conditions of life; or by having adapted them during long-past periods of time: the adaptations being aided in some cases by use and disuse, being slightly affected by the direct action of the external conditions of life, and being in all cases subjected to the several laws of growth. Hence, in fact, the law of the Conditions of Existence is the higher law; as it includes, through the inheritance of former adaptations, that of Unity of Type. (Darwin 1859. 206.)

Note, what Darwin always believed, that natural selection is not the sole causal force for change. Darwin always accepted, what we now think is wrong, that the inheritance of acquired characteristics – usually known as "Lamarckism," after the use of it by the French evolutionist, Jean Baptiste de Lamarck – plays a role in the evolutionary story. "Wax of Ear, bitter perhaps to prevent insects lodging there, now these exquisite adaptations can hardly be accounted for by my method of breeding there must be some cor[r]elation, but the whole mechanism is so beautiful" (Darwin 1987. C 174). It is just that, alone, Lamarckism is not adequate.

VIII. DARWIN AND RELIGION

So much for Darwin's *Origin*. Teleology without tears. Final cause accepted and highlighted, but under the machine root metaphor. Before we move on, it would be ungracious not to acknowledge that all who write on the topic of Darwin and teleology are hugely indebted to a 1993 article, "Darwin was a teleologist," in *Biology and Philosophy* (a journal of which I was the founding editor), by James Lennox. He shows unambiguously that Darwin was a teleologist (for reasons given in the last section); additionally, he refutes those – for example, the biologist Michael Ghiselin – who argued that there was no such teleology, that Darwin had taken it out of biology, and that Darwin was consciously aware of what he was doing. Ghiselin, for instance, referred to the underlying teleology of Darwin's next book after the *Origin – The Various Contrivances by Which Orchids are Fertilised by Insects* (1862) – as a "metaphysical satire" (Ghiselin 1969. 135). Lennox shows not only how mistaken an interpretation that is, but that back when Darwin became an evolutionist and discovered natural selection, he was already facing the fact that, although he had now taken God out of the equation,

the same could not be said of "final causes," teleology. In an unpublished comment (written in 1838) in the margin of a book he was then reading – *Proofs and Illustrations of the Attributes of God*, by John Macculloch – Darwin wrote: "The Final Cause of innumerable eggs is explained by Malthus – (is it anomaly in me to talk of Final Causes: consider this! –) consider these barren Virgins" (Darwin 1987. 637). Remember that, in the *Origin*, Darwin was still worrying about those eggs! The reference to "barren virgins" refers to the already-encountered description of final causes by Francis Bacon. Darwin would have picked it up from William Whewell's natural-theology-promoting *Bridgewater Treatise* (1833. 355–356). Clearly, anomaly or not, Darwin decided that he could legitimately go on using the term.

Ask now some questions that arise from the discussion. First, what about religion? Does Darwinian evolutionary theory, with natural selection as its central mechanism, refute God, specifically the Christian God? If so, it would have been a surprise to Darwin! Towards the end of the *Origin*, all six editions (last in 1872), Darwin affirms his belief in the possibility of religious acceptance. Indeed, his position makes it easier.

> Authors of the highest eminence seem to be fully satisfied with the view that each species has been independently created. To my mind it accords better with what we know of the laws impressed on matter by the Creator, that the production and extinction of the past and present inhabitants of the world should have been due to secondary causes, like those determining the birth and death of the individual. When I view all beings not as special creations, but as the lineal descendants of some few beings which lived long before the first bed of the Silurian system was deposited, they seem to me to become ennobled. (488–489)

It is true that Darwin is pushing one to the God of deism – He works through and only through unbroken law – rather than the God of theism – God works through miracles. Darwin was hardly the first to go this way. Privately, Newton was a deist. Moreover, by the nineteenth century, many, independently of science, were starting to make miracles law-bound. The Marriage at Cana, where Jesus turns water into wine, is best understood, not as conjuring, but as a tale where Jesus so moved the party-giver that he voluntarily opened up his cellars and brought out his best wine. Many today, indeed, would say that calling for divine intervention is precisely to miss the meaning of the event. In the years of my childhood, the years after the Second World War, the British considered Dunkirk in 1940, when the British Army escaped across the Channel, to be a miracle. They were able to regroup and continue the fight against Hitler. God did not make it easy for them; He made it possible for them. If you had asked the average Brit whether God did it through special intervention or through blind law, they would have looked at you as though you were queer in the head.

Or making a somewhat inappropriate joke. What mattered was the meaning not the cause (Ruse 2001).

Clearly Darwin is aiding the cause of law-bound explanations, whether Christians like this or not. Is Darwin truly setting us on the road towards disbelief? After all, despite what he said in the *Origin*, by about 1870 he had become what Thomas Henry Huxley called an "agnostic." Neither believer nor non-believer. However, in common with just about every Victorian agnostic, and as the nineteenth century drew to a close there were many of them, Darwin's chief gripe against Christianity was theological. In his autobiography written about 1876, he wrote:

> I gradually came to disbelieve in Christianity as a divine revelation. The fact that many false religions have spread over large portions of the earth like wild-fire had some weight with me. Beautiful as is the morality of the New Testament, it can hardly be denied that its perfection depends in part on the interpretation which we now put on metaphors and allegories.
>
> But I was very unwilling to give up my belief; – I feel sure of this for I can well remember often and often inventing day-dreams of old letters between distinguished Romans and manuscripts being discovered at Pompeii or elsewhere which confirmed in the most striking manner all that was written in the Gospels. But I found it more and more difficult, with free scope given to my imagination, to invent evidence which would suffice to convince me. Thus disbelief crept over me at a very slow rate, but was at last complete. The rate was so slow that I felt no distress, and have never since doubted even for a single second that my conclusion was correct. I can indeed hardly see how anyone ought to wish Christianity to be true; for if so the plain language of the text seems to show that the men who do not believe, and this would include my Father, Brother and almost all my best friends, will be everlastingly punished.
> And this is a damnable doctrine. (Darwin 1958. 86–87.)

Darwin was an agnostic. Yet, an agnostic of a recognizable kind. For some, who call themselves "agnostic," this is really a way of saying: "I couldn't care less. I really find the whole topic rather boring." (My wife falls into this category.) For others, in its way, agnosticism is as dynamic as full-blooded belief. (I fall into this category!) The eminent population geneticist J. B. S. Haldane wrote: "Not only is the world queerer than we think it is. It is queerer than we could think it is." This is not a man who has shelved the problem. The ultimate meaning of things is a challenging mystery. This was Darwin's position. Traditional Christianity is false and morally offensive. Deism, the long-held position, is truly knocked sideways by the law-bound process of natural selection. And yet…..? Towards the end of his life, to a correspondent who had just sent him a book on issues to do with science and religion, Darwin wrote:

> You would not probably expect anyone fully to agree with you on so many abstruse subjects; and there are some points in your book which I cannot digest. The chief one is that the existence of so-called natural laws implies purpose. I cannot see this. Not to mention that many expect that the several great laws will some day be found to follow inevitably from some one single law, yet taking the laws as we now know them, and look at the moon, what the law of gravitation – and no doubt of the conservation of energy – of the atomic theory &c. &c. hold good, and I cannot see that there is then necessarily any purpose. Would there be purpose if the lowest organisms alone destitute of consciousness existed in the moon? (Letter to William Graham, July 3, 1881. Darwin Correspondence Project, letter:13230.)[3]

Darwin continues:

> Nevertheless you have expressed my inward conviction, though far more vividly and clearly than I could have done, that the Universe is not the result of chance. But then with me the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey's mind, if there are any convictions in such a mind?

An agnostic indeed!

Should Darwin have gone all the way to atheism? This seems to be the assumption of many. The title of Sam Harris's book, *The End of Faith*, tells the tale. He states flatly that "the truth is that religious faith is simply unjustified belief in matters of ultimate concern – specifically in propositions that promise some mechanism by which human life can be spared the ravages of time and death. Faith is what credulity becomes when it finally achieves escape velocity from the constraints of terrestrial discourse – constraints like reasonableness, internal coherence, civility, and candor" (Harris 2004. 65).

Go back to Hume. He then was caught on the argument to the best explanation. Now, a law-bound explanation of design is no longer impossible. The way was open to Hume to declare for atheism. Whether he would have done is another matter. Whether Darwin would have forced him to become an atheist is up for doubt. Richard Dawkins (1986) has said "Darwin made it possible to be an intellectually fulfilled atheist." There is certainly no compulsion to be an atheist. Indeed, you can go on believing fully as a Christian, although you might now be more inclined to put your money on revealed religion rather than natural religion. This is a stance taken independently by many Christians in the last two centuries. Inspired particularly by Søren Kierkegaard, the feeling is that faith is undercut if it is backed up by reason. Faith is no longer courageous, if it

[3] https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-13230.xml#Lfoot_f2

is no longer a leap into the absurd. Many would not go this far but would agree – with the traditional position of St Thomas – that faith must come first. This was certainly the stance of the great nineteenth-century theologian John Henry Newman. "I believe in design because I believe in God; not in a God because I see design" (Newman 1973. 97). As a Christian, one believes on faith all about the Christian God, and then one fleshes this out by looking at the world and using reason. After all, that is what being made in the image of God is all about.

Whatever you may think about the argument from design, this does not exhaust natural theology. There are other proofs for the existence of God, and there are still arguments against the existence of God. We have seen reason to think that the argument from miracles is perhaps less convincing than formerly – Darwin's theory does not deny miracles, but it certainly starts to make divine intervention less pressing. Other arguments – the ontological argument and the causal argument, for instance – have to be considered independently, on their merits. The biggest argument against God is the traditional argument from evil. An all-powerful, all-knowing God, all-loving God would not let evil exist. This powerful passage is from *The Brothers Karamazov*:

> "Tell me yourself, I challenge your answer. Imagine that you are creating a fabric of human destiny with the object of making men happy in the end, giving them peace and rest at last, but that it was essential and inevitable to torture to death only one tiny creature – that baby beating its breast with its fist, for instance – and to found that edifice on its unavenged tears, would you consent to be the architect on those conditions? Tell me, and tell the truth."
> "No, I wouldn't consent," said Alyosha softly.
> (Dostoevsky 1879. ch. 4.)

There are some things an all-loving God would not allow, not even for the eternal salvation of every human being, past and present. And, remember, this was written before the Holocaust.

Traditionally the response to the problem of evil divides it into two: natural evil and moral evil (Ruse 2001). Natural evil focusses on natural mishaps like the Lisbon earthquake; less dramatic, like the painful, incurable cancer of a small child (Davies and Ruse 2021). Moral evil focuses on free will. It is better that Heinrich Himmler had free will, than not, even though it did lead to the Final Solution and the death of six million Jews. Interestingly, Darwinism has been taken as relevant to both natural and moral evil. Even more interestingly – perhaps "paradoxically" is a better word – Darwin has been taken as supportive of the two approaches. In the case of natural evil, it is Richard Dawkins (1983) of all people who has made the point that natural selection clearly leads to pain and suffering. That is what a struggle for existence is all about. Darwin wrote to Asa Gray on the subject. "I cannot persuade myself that a beneficent & omnipotent

God would have designedly created the Ichneumonidæ with the express intention of their feeding within the living bodies of caterpillars, or that a cat should play with mice." (Letter to Asa Gray, *22*, May 1860, in Darwin 1985/8. *224*.) Dawkins stresses what we have been stressing, namely the design-like nature of organisms, and argues that the only way that such design-like organisms could have been created is through natural selection. Lamarckism, the inheritance of acquired characteristics, is empirically false, saltations (jumps from one form to another) are inadequate – they just lead to randomness – and there really is no other game in town. So, natural evil is an inevitable consequence of getting organisms naturally, and this includes humans.

Moral evil, depending on free will, raises the question of the plausibility of the free will. Darwinism stresses that nature is law-bound. Doesn't this preclude freedom of choice? Calvinists are right. Everything is predestined. Free will is only possible if we can escape law, and we cannot. In response, philosophers distinguish between two takes on the free will problem. Libertarianism, which has nothing to do with the political philosophy of Ayn Rand, says we can escape laws. Kant thought this possible. Compatibilism, free will can occur only with a law frame. Hume thought this, probably reflecting the Calvinist background of Protestant Scots. In America, Jonathan Edwards endorsed it. In support of their position, compatibilists argue that absence of laws does not imply freedom. It implies craziness. If the late Queen had taken off all her clothes before she appeared on the balcony of Buckingham Palace, we would not applaud her actions, but worry about her mental health. All training is designed, not to preclude freedom, but the very opposite: the freedom to make reasoned choices and not to act on blind prejudice.

Understood in this light, the Darwinian is clearly going to be a compatibilist. Now add a nice point. Evolutionists distinguish between r reproductive strategies and K reproductive strategies (MacArthur and Wilson 1967). The former, r strategies, puts the emphasis on having lots of offspring but little parental care. Herrings. The latter, K strategies, puts the emphasis of few offspring but much parental care. Primates. The former strategy makes sense when conditions fluctuate. Famine or feast. The r strategy can take full advantage of good times and these more than balance bad times. The K strategy makes sense when conditions are stable. You can rely on a steady background and take time raising offspring. Humans, obviously, are the supreme K strategists. Think of the time it takes for our offspring to mature. The r strategist has little need of free will. If a rain shower washes away a crowd of ants, too bad. Rather than putting effort into raising far fewer who might react to the shower and try to escape, the Queen is better off producing many more to take their place. Humans cannot afford to lose offspring every time it rains. So, we need to have a dimension of freedom. If it starts to rain, stop shopping and go to Starbucks for a latte, until it is over. We are like Mars Rover (Dennett 1984). It is completely governed by law, but it

does not have to wait for instructions from Earth every time it meets an obstacle. A rock is in the way? Go around it, rather than come to grief trying to ride up the side. In other words, on both fronts, Darwinism is supportive of traditional answers to the problem of evil. This is not to say that they are now adequate. I doubt anything like this is going to move Dostoevsky's Alyosha. I suspect most people are supremely unworried as to whether Heinrich Himmler is or is not like Mars Rover. He was grotesquely inhumane and no excuse about the value of his free will is going to affect that judgment. Enough said.

IX. FINE TUNING

Or is it enough? Have we perhaps sold design short? Drawing to a conclusion to this essay, I want to look at two groups who think the discussion is ending too quickly. There is more to design than the eviscerated machine-metaphor analysis that Darwin offers us. Some think this opens the way back to the Christian God; others think, perhaps God but not the traditional Christian version; and yet others think, maybe no God at all. I shall look first at a group that strikes me as putting enthusiasm and wish fulfillment above critical thinking. Then a group that offers a much more interesting challenge to the Darwinian position presented in this essay. They may or may not be right, but they should be taken seriously.

To start with the first group, its members champion design, not in biology, but in physics. This is the so-called "fine-tuning" argument, that argues the basic constants of the universe are not random, but carefully thought out and chosen, else life and much else would be impossible (Friederich 2001). Hence, God makes a comeback. Why should we think the universe is fine-tuned? Several physical arguments are offered, all along the line of "if this had not been exactly as it is, that would not have happened, and so no life would have been possible." What would be an example? The carbon atom is a popular choice (Weinberg 1999). In the early stages of the universe there were no carbon atoms. At that point, everything was just hydrogen and helium. For carbon to be produced, we need three helium nuclei. Normally, even with the right ingredients nothing happens because the energy of carbon is way below that of three helium nuclei – as things normally are, the nuclei could not come together and stay that way. They are too hyped up as it were. Fortuitously, however, there is a variant, radio-active form of carbon. It has just the higher energy that is needed and so everything works out perfectly – this energy of the radio-active form is precisely that needed to make carbon. Anything a little more, it would not work. Anything a little less, it would not work. The actual energy level is right on target. Like Goldilocks' third try at the Three Bears dishes of porridge, it is just fine. But before you get all excited and think that nature is not just fine but fine-tuned,

the very skeptical physics Nobel Laureate Steven Weinberg asks us to keep questioning. How do you get the three helium nuclei in the first place? They come together in a two-part process. First, two of them combine to make beryllium. Only then is the third is added to make carbon. It turns out that looking at things from this perspective there is a lot more room for flexibility – there is a wider range of energy levels that would let these processes move forward. There is thus no unique possible energy needed to make carbon. All in all, therefore, perhaps things are not so tightly designed.

The trouble with the arguments in favor of fine-tuning is that we are just working from ourselves – from the world we know – and putting probabilities on things is such guesswork. Think of a number, double it, and the answer you want is a half. The fine-tuning enthusiasts start from premises no one would deny. Of course, we humans could not function on a planet where, because it is bigger, the gravitational attraction is (let us say) twice as strong. As we are constituted now, the strain on our limbs and our internal organs like the heart would lead to early death. But then the fine tuners go astray by assuming that this is all there is to be said on the subject. This is a mistake. If we were on a bigger planet, then natural selection would have made us so that we could live there. We might, for instance, have evolved with elephantine-sized legs. Or more plausibly, perhaps like the whales we could have spent most of our time in the water where we would weigh that much less, and so presumably we would have adaptations like dolphins for living an aquatic life, our hearts, and lungs and (obviously most important) brains could be very human-like. I am not sure that advanced civilization is beyond mermen and mermaids. And this is all before you start to think of the trendy new notion of "multiverses" (Ellis 2011). Perhaps our universe is just one of an infinite number, some of which work, some of which don't, some of which support life, some of which don't. We are right back to winning the lottery without any fraud behind our success. We couldn't buy the Mercedes if we hadn't won it, but winning it was no miracle.

X. ROMANTICISM

Turn now to the second, more-interesting challenge to the Darwinian analysis. By the end of the eighteenth century, with the failure of mechanism to explain organisms, there were those who started to champion the organicist metaphor, thinking that in the Scientific Revolution it had been too quickly discarded (Cunningham and Jardine 1990; Richards 2003). These "Romantics", as they were called, included the poet Johann Wolfgang von Goethe, the anatomist Lorenz Oken, and above all the philosopher Friedrich Schelling (Knight 1990). As a teenager, Schelling had written a sixty-page essay on the *Timaeus*. It had a lasting influence. "The key to the explanation of the entirety of the Platon-

ic philosophy is noticing that Plato everywhere carries the subjective over to the objective" (Schelling 1833. 212). Schelling saw the world in organic terms, meaning that he thought there is value to be found out in the world, it is not just ascribed by us to value-free machines.

> Even in mere organized matter there is life, but a life of a more restricted kind. This idea is so old, and has hitherto persisted so constantly in the most varied forms, right up to the present day – (already in the most ancient times it was believed that the whole world was pervaded by an animating principle, called the world-soul, and the later period of Leibniz gave every plant its soul) – that one may very well surmise from the beginning that there must be some reason latent in the human mind itself for this natural belief. (Schelling 1803. 35.)

The world is something that produces itself, has its developing powers inside, as an unfurling organism is driven by forces within rather than without. One goes from the simple to the complex, from the undifferentiated to the highly differentiated. "Nature should be Mind made visible, Mind the invisible nature. Here then, in the absolute identity of Mind in us and Nature outside us, the problem of the possibility of a Nature external to us must be resolved. The final goal of our further research is, therefore, this idea of Nature; if we succeed in attaining this, we can also be certain to have dealt satisfactorily with that Problem" (42). Schelling saw the world in constant motion. And we humans come at the top. "It is One force, One interplay and weaving, One drive and impulsion to ever higher life" (Schelling: *Proteus of Nature*, 1800, in Morgan 1990. 35). Note that we have now an extra dimension to purpose. There is the purpose as exhibited by adaptation. The purpose of the teeth is to bite off and chew one's food. And now, more explicitly, we have purpose in a historical mode. Things don't just change, they change in order to point us ever closer to the apotheosis of the historical story. Humankind! Progress!

Note the relevance of all of this to the theme of this essay. For the Darwinian, design is a product of blind law. It is brought on by the external force of natural selection. Design in itself has no absolute value. It is neither good nor bad. It is we who make the judgment. The eye of humans is a good thing for us. The fang of the snake is a bad thing for us, although it might well be a very good thing for the snake. Under the organicist model, the design flows naturally from within. The flower grows naturally, first a bud and then an opening in all its splendor and functioning to attract pollinating insects. For the Platonist, the design is Design, produced by an intelligence. For the Aristotelian it is something that emerges from the natural value-laden laws of nature. They are infused with soul in some sense.

Where this leaves someone like Schelling is a matter for inquiry. Someone like him could be a Christian but equally they might be a non-believer, simply

thinking that the value-laden world is the way things are. Certainly, for Schelling, God is within the organicism circle, developing and of great value. "God is himself bound to nature through freely willed love; he does not require her and yet will not exist without her. For love is not the result of two beings requiring one another, but it occurs when each could exist for itself,… yet where neither can exist morally without the other" (Richards 2003. 146). There is a shift from traditional Protestant theology. God traditionally is thought not to want anything from us. In the words of Martin Luther: "a Christian lives not in himself, but in Christ and in his neighbor. Otherwise he is not a Christian. He lives in Christ through faith, in his neighbor through love. By faith he is caught up beyond himself into God. By love he descends beneath himself into his neighbor" (Luther 1970. 309). Schelling's idealism, his organicism, implying his holism for the plant develops as a whole and not in parts, means that God is interacting with us. He is not the eternal, separate entity posited by Augustine and others.

Much influenced by Schelling was Darwin's contemporary Herbert Spencer (Ruse 2021). An evolutionist, he thought less in terms of natural selection and more in terms of Lamarckian processes, the inheritance of acquired characteristics (Spencer 1852; Richards 1987). He was a holist, thinking societies are like organisms (Spencer 1860). And he was a fanatical progressionist.

> This law of organic progress is the law of all progress. Whether it be in the development of the Earth, in the development of Life upon its surface, in the development of Society, of Government, of Manufactures, of Commerce, of Language, Literature, Science, Art, this same evolution of the simple into the complex, through successive differentiations, holds throughout. (Spencer 1857. 245.)

He explained that the English language is more complex and hence above all others. Expanding on this, grabbing ideas from physics, Spencer suggested that external forces cause things to get out of equilibrium, then as they strive to reachieve equilibrium, they rise higher. History therefore is a series of stages, going from one stable level to another (higher) one. "Dynamic equilibrium" (Spencer 1862).

Following Spencer came the French philosopher Henri Bergson, author of *L'évolution créatrice*, published in 1907 (English translation 1911), champion of the neo-Aristotelian life force, the *élan vital* – hence, better known as a "vitalist" rather than the more comprehensive "organicist". The philosophy is the same and is derivative: deeply Aristotelian, including the importance of final cause. "The 'vital principle' may indeed not explain much, but it is at least a sort of label affixed to our ignorance, so as to remind us of this occasionally, while mechanism invites us to ignore that ignorance" (Bergson 1911. 42). Expectedly, vitalism speaks to "internal finality." With predictable conclusions: "not only does consciousness appear as the motive principle of evolution, but also, among

conscious beings themselves, man comes to occupy a privileged place. Between him and the animals the difference is no longer one of degree, but of kind" (Bergson 2011. 34). More than this even: "in the last analysis, man might be considered the reason for the existence of the entire organization of life on our planet" (35).

A little later, crossing the Atlantic, we encounter the transferred Englishman, Alfred North Whitehead (1926). The world has value, in some sense it is living, and so naturally one thinks of mind as being all-pervasive. "The doctrine that I am maintaining is that neither physical nature nor life can be understood unless we fuse them together as essential factors in the composition of 'really real' things whose interconnections and individual characters constitute the universe" (Whitehead 1938. 205). Continuing: "this sharp division between mentality and nature has no ground in our fundamental observation. We find ourselves living within nature." Hence: "I conclude that we should conceive mental operations as among the factors which make up the constitution of nature" (214). It is the perceived unacceptability of the traditional God of Christianity, eternal and unchanging, that is the raison d'être for Whitehead's approach to the God problem, developed as it was into so-called "Process Theology." Whitehead and his followers wanted nothing to do with a God who is unmoved – could not be moved because He is eternal and unchanging – by the death of Anne Frank in Bergen-Belsen. In any case, as an out-and-out follower of Schelling, on the one hand Whitehead took the inherent change of organicism as all-important, and, on the other hand, was totally committed to a God in the world rather than a God who is in some sense logically separate. Remember: "Nature should be Mind made visible, Mind the invisible nature. Here then, in the absolute identity of Mind in us and Nature outside us, the problem of the possibility of a Nature external to us must be resolved" (Schelling 1803. 42). Whitehead writes:

> The vicious separation of the flux from the permanence leads to the concept of an entirely static God, with eminent reality, in relation to an entirely fluent world, with deficient reality. But if the opposites, static and fluent, have once been so explained as separately to characterize diverse actualities, the interplay between the thing which is static and the things which are fluent involves contradiction at every step in its explanation. (Whitehead 1929. 346)

Continuing:

> The final summary can only be expressed in terms of a group of antitheses, whose apparent self-contradictions depend on neglect of the diverse categories of existence. In each antithesis there is a shift of meaning which converts the opposition into a contrast.

It is as true to say that God is permanent and the World fluent, as that the World is permanent and God is fluent. Whitehead's God is a God who evolves with us, working with us to achieve progress, a better world.

Moving to the present and to science, through the mentors he had as a graduate student at Harvard, the eminent evolutionist Edward O. Wilson was deeply influenced by Spencer. In his major work on the evolution of social behavior, *Sociobiology: The New Synthesis*, Wilson tells us that of all animals: "Four groups occupy pinnacles high above the others: the colonial invertebrates, the social insects, the nonhuman mammals, and man" (Wilson 1975. 379). He continues: "Human beings remain essentially vertebrate in their social structure. But they have carried it to a level of complexity so high as to constitute a distinct, fourth pinnacle of social evolution" (380). He concludes by speaking of humans as having "unique qualities of their own." He now launches at length into showing us how humans have crossed over and mounted the "fourth pinnacle" (382) – the "culminating mystery of all biology" (382). All this, as Wilson makes clear in subsequent writings, is very much part of the general picture. "The overall average across the history of life has moved from the simple and few to the more complex and numerous. During the past billion years, animals as a whole evolved upward in body size, feeding and defensive techniques, brain and behavioral complexity, social organization, and precision of environmental control – in each case farther from the nonliving state than their simpler antecedents did" (Wilson 1992. 187). Wilson talks of selection, but it is not the traditional selection of Darwinism, where adaptations are always for the individual. Wilson sees selection acting for groups and hence there is a kind of integration, holism, about the nature of species (Wilson and Wilson 2007). If this isn't an organicist picture of life's history, it is hard to know what would be. One doubts that Wilson has even heard of Friedrich Schelling, let alone read him, but the tradition lives on (Gibson 2013).

There is today a vibrant group of evolutionary biologists who declare for organicism – the "New Biologists" (Laland et al 2014, 2015; Bateson et al 2017). But, to conclude this brief survey, turn to the philosophers, for there too we find much enthusiasm. British philosopher John Dupré is blunt. "There are powerful reasons for thinking that emancipation from the mechanistic paradigm is a precondition for true insight into the nature of biological processes" (Dupré 2012. 83). We learn that, at best, natural selection does little. "Where does adaptive change come from? A trivial but sometimes obfuscated point is that it never comes from natural selection." Continuing: "Selection cannot occur unless some other process provides alternatives to select from. It follows that any thesis about the power of natural selection to generate change implicitly presupposes a thesis about a process or processes that generate selectable change." The reader will not be surprised to learn that "our forms of consciousness of which we are capable, are very different from those of other terrestrial animals." Likewise with

human culture. It "involves the articulation and synchronization of a variety of roles and functions that is different in kind from anything else in our experience." Adding: "our forms of consciousness of which we are capable, are very different from those of other terrestrial animals."

Fellow philosopher Jerry Fodor (2007) feels much the same way. Of the correct evolutionary picture, we learn: "The slogan is the evolution of ontogenies. In other words, the whole process of development, from the fertilized egg to the adult, modifies the phenotypic effects of genotypic changes, and thus 'filters' the genotypic options that ecological variables ever have a chance to select from" (Fodor and Piattelli-Palmarini 2010. 27). And that of course is precisely what the Romantics claim. Look at the development of the individual – the growth of the chimpanzee – you have the answer to the evolution of the group – the evolution of the primates.

Finally, another fellow philosopher, Thomas Nagel (2010), stresses that it is precisely the problem of design that makes him turn from the Darwinian, mechanical explanation. He speculates that possibly "there are natural teleological laws governing the development of organization over time, in addition to laws of the familiar kind governing the behavior of the elements." He allows that: "This is a throwback to the Aristotelian conception of nature, banished from the scene at the birth of modern science. But I have been persuaded that the idea of teleological laws is coherent, and quite different from the intentions of a purposive being who produces the means to his ends by choice. In spite of the exclusion of teleology from contemporary science, it certainly shouldn't be ruled out a priori" (22). One should add that Nagel is an avowed atheist, so a Platonic option is not really open. As he himself says, Nagel is looking more for "natural teleological laws."

XI. PROGRESS

Note something of importance. Dupré particularly has more to his thinking about design (real or apparent) than adaptation, such as the working of the hand or the eye. He is also thinking historically. He sees purpose in the course of evolution. Monad to man (Ruse 2017). Dupré puts humans above other organisms. Evolution for him is *progressive*. In this belief, as we have seen, Dupré belongs to a long tradition. Above all the "Romantics." Goethe and Schelling. Then, their English disciple Herbert Spencer. And so down to the present and to Edward O Wilson. Was Darwin indifferent to all this? Even if he showed that teleology at the individual level, adaptation, could be explained within the mechanical paradigm, did he quietly avoid teleology at the historical level? Some did this for Darwin. The German evolutionist Ernst Haeckel (1834–1919) claimed – and it does seem in respects that he genuinely thought – he was a great disciple

of Darwin. But if one looks carefully at his writing and theorizing, he sounds much more Romantic – organismic – than Darwinian – mechanistic. This came naturally to one who cut his scientific teeth on embryological studies – the area of biology focusing on the development, irrespective of outside forces, of the fertilized egg to the full-grown adult. This hints – more than hints – that biological development, change of any kind, is going to be fueled from within, as it were, rather than from without, which latter is precisely the way that the force of natural selection works. This belief was confirmed by Haeckel's championing of the "biogenetic law": "ontogeny recapitulates phylogeny." As the individual organism develops it is precisely mimicking the way that the group develops. Confirming Haeckel's debts to Romanticism, in looking at the many phylogenies that he drew – he was a talented illustrator – we inevitably see progress, usually progress to human beings. Haeckel's vision of evolution was value-laden in a way we have just seen Darwin explicitly eschewing. But what then was Darwin's response to the challenge of progress? Above all, he strove to keep values out of his science. Most particularly in repudiating claims about progress and humans at the top. From the beginning of his thinking about evolution, as soon as he discovered natural selection, he was arguing that it gives no guarantee of progress. What else would one expect from someone so hugely within Lyell's uniformitarian orbit? "The enormous number of animals in the world depends of their varied structure & complexity. – hence as the forms became complicated, they opened fresh means of adding to their complexity. – but yet there is no necessary tendency in the simple animals to become complicated although all perhaps will have done so from the new relations caused by the advancing complexity of others" (E97, written in January 1839). On the flyleaf of his copy of a pre-*Origin* evolutionary *Vestiges of the Natural History of Creation*, he cautioned himself never to use the terms "higher" and "lower."

Darwin kept on worrying about this issue. In the first edition of the *Origin*, 1859, he does allow a kind of progressive odor to the fossil record, but it is hardly an enthusiastic endorsement. In the third edition of the *Origin*, 1861, just two years after the first edition, he added several new paragraphs on the topic. He basically repeated the sentiment in his notebooks about organization leading to highness. "If we look at the differentiation and specialisation of the several organs of each being when adult (and this will include the advancement of the brain for intellectual purposes) as the best standard of highness of organisation, natural selection clearly leads towards highness;…" But then, later – in this same edition – he qualified what he had said to be virtually vacuous:

> To attempt to compare in the scale of highness members of distinct types seems hopeless: who will decide whether a cuttlefish be higher than a bee – that insect which the great Von Baer believed to be "in fact more highly organised than a fish, although upon another type"? In the complex struggle for life it is quite credible that

> crustaceans, for instance, not very high in their own class, might beat the cephalopods or highest molluscs; and such crustaceans, though not highly developed, would stand very high in the scale of invertebrate animals if judged by the most decisive of all trials – the law of battle.

Darwin kept emphasizing the underlying sentiment even after the *Descent* was published. To the American evolutionist Alphaeus Hyatt he wrote: "After long reflection I cannot avoid the conviction that no innate tendency to progressive development exists, as is now held by so many able naturalists, & perhaps by yourself" (Letter, December 4, 1872).

By the mid-nineteenth century, mechanism was proving its worth again and again. Darwin's *Origin* apparently proved this. He set out to give the biological equivalent of Newtonian mechanics, the final stage of the effort to show that the world could be explained by scientific theories guided by the machine root metaphor. He accepted teleology. Then, he offered an account of teleology that fell under this metaphor. It is true that there were/are those who thought/think that one can remain a mechanist and yet believe in progress. There is no contradiction in you making the judgment that humans are above all other animals. It is just that this is your judgment and not something you derive from Darwin's theory of evolution through natural selection. In the immortal words of the paleontologist Jack Sepkoski: "I see intelligence as just one of a variety of adaptations among tetrapods for survival. Running fast in a herd while being as dumb as shit, I think, is a very good adaptation for survival" (Ruse 1996. 486).

Candor demands that one admit there are those, committed Darwinians, who nevertheless think that the theory supports notions of progress. Richard Dawkins is one such person. "Directionalist common sense surely wins on the very long time scale: once there was only blue-green slime and now there are sharp-eyed metazoan" (Dawkins 1986. 38). He finds the key in "arms races." As one who embraced computer technology early and enthusiastically, perhaps expectedly Dawkins notes that, more and more, today's arms races rely on computer technology rather than brute power, and – in the animal world – he finds this translated into ever-bigger and more efficient brains. No need to hold your breath about who has won. Dawkins invokes a notion known as an animal's EQ, standing for "encephalization quotient" (Dawkins 1986. 39). This is a kind of cross-species measure of IQ that takes into account the amount of brain power needed simply to get an organism to function (whales require much bigger brains than shrews because they need more computing power to get their bigger bodies to function), and that then scales according to the surplus left over. Dawkins writes: "The fact that humans have an EQ of 7 and hippos an EQ of 0.3 may not literally mean that humans are 23 times as clever as hippos! But the EQ as measured is probably telling us something about how much 'computing power' an animal probably has in its head, over and above the irreducible amount

of computing power needed for the routine running of its large or small body." As always, it is the analogy with human progress that is the key. Computer evolution in human technology is enormously rapid and unmistakably progressive. It comes about through at least partly a kind of hardware/software coevolution. Advances in hardware are in step with advances in software.

> There is also software/software coevolution. Advances in software made possible not only improvements in short-term computational efficiency – although they certainly do that – they also make possible further advances in the evolution of the software. So the first point is just the sheer adaptedness of the advances of software make for efficient computing. The second point is the progressive thing. The advances of software, open the door – again, I wouldn't mind using the word "floodgates" in some instances – open the floodgates to further advances in software. (Ruse 1996. 469.)

He adds, "I was trying to suggest, by my analogy of software/software coevolution, in brain evolution that these may have been advances that will come under the heading of the evolution of evolvability in the evolution of intelligence."

Let us leave things at that. Critics are going to be less than enthused by computer-fueled advance. Anyone who thinks that the development of technology will always spell progress is an optimist indeed. Is it really the case that in the next twenty thousand years no mad fools will find a way to destroy us all? All one can say is that, even for mechanists, progress can be an elusive and much-desired vision. And this apart from the fact that there were (and still are) those who regretted the demise of the organic root metaphor. Somehow there was a feeling that something of value had been lost. Something of spiritual value, without necessarily being overtly Christian.

XII. CONCLUSION

As always, Hume had the measure of things. "In subjects adapted to the narrow compass of human reason, there is commonly but one determination, which carries probability or conviction with it; and to a man of sound judgement, all other suppositions, but that one, appear entirely absurd and chimerical" (Hume 1779. 81). The problem is that men of "sound judgement" so often come to different conclusions. Whewell thought he was right. Darwin thought he was right. The Romantics, Schelling to Nagel, think that they are right. I am not sure that it is my job here to make a decision. I think we can fairly say that Darwin had the measure of the traditional organicists, from Plato through to Whewell. He explained design as a matter of blind laws, eternally in motion. At the same time, he explained the problems for traditional design, such as the homologies between organisms. We have just seen, however, that organicism

may have been floored. A vigorous group argue that it is not out. One should add that Darwinians argue with no less vigor that organicism is still not adequate. Responding to Dupre's musings, Chicago evolutionary biologist Jerry Coyne replies: "We do not need a new philosophical framework for evolution, much as Dupré wants one. Traditional reductionist views are still valid and yielding valid insights (what is microRNA other than a 'bottom-up' phenomenon that regulates genes?)." He adds: "As an evolutionary biologist – which Dupré is not – I think I'd know if my field was in crisis. Yet I haven't heard any recent lamentations from my colleagues" (Coyne 2012).

One might feel that Coyne is just stating his position rather than arguing for it. I suspect he would return the challenge to the critics. Highly regarded today is the work of Peter and Rosemary Grant on the evolution of the finches on the Galapagos Archipelago (Grant and Grant 2014). Their work is so highly regarded that a Pulitzer Prize winning book was published about their work (Weiner 1994). Tell us, he would say, what is inadequate about this science. And with this rhetorical question, I will leave matters there. What comes next is an exercise for the reader!

REFERENCES

Aquinas, St. Thomas. 1981. *Summa Theologica*. Trans. the Fathers of the English Dominican Province. London, Christian Classics.

Aristotle 1984a. De Generatione de Animalium. In Jonathan Barnes (Ed.) *The Complete Works of Aristotle*. Princeton, Princeton University Press. 1111–1218.

Aristotle 1984b. De Partibus Animalium. In Jonathan Barnes (Ed.) *The Complete Works of Aristotle*. Princeton, Princeton University Press. 1087–1110.

Bateson, Patrick – Nancy Cartwright – John Dupré – Kevin Laland – Denis Noble 2017. New Trends in Evolutionary Biology: Biological, Philosophical and Social Science Perspectives. Special issue of *Interface Focus*. https://doi.org/10.1098/rsfs.2017.0051

Bergson, Henri 1907. *L'évolution créatrice*. Paris, Alcan.

Bergson, Henri 1911. *Creative Evolution*. New York, Holt.

Boyle, Robert 1688/1966. A Disquisition about the Final Causes of Natural Things. In T. Birch (Ed.) *The Works of Robert Boyle*. 5. Hildesheim, Georg Olms. 392–444.

Boyle, Robert 1996. *A Free Enquiry into the Vulgarly Received Notion of Nature*. Ed. E. B. Davis – M. Hunter. Cambridge, Cambridge University Press.

Brewster, David 1854. *More Worlds than One: The Creed of the Philosopher and the Hope of the Christian*. London, Camden Hotten.

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

Coyne, Jerry A. 2012. Another Philosopher Proclaims a Nonexistent "Crisis" in Evolutionary Biology. *Why Evolution Is True Blog*. https://whyevolutionistrue.com/2012-09/07/another-philosopher-proclaims-a-nonexistent-crisis-in-evolutionary-biology/

Cunningham, Andrew – Nicholas Jardine (Ed.) 1990. *Romanticism and the Sciences*. Cambridge, Cambridge University Press.

Darwin, Charles 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London, John Murray.

Darwin, Charles 1862. *On the Various Contrivances by which British and Foreign Orchids are Fertilized by Insects, and On the Good Effects of Intercrossing.* London, John Murray.
Darwin, Charles 1958. *The Autobiography of Charles Darwin, 1809-1882.* Ed. Nora Barlow. London, Collins.
Darwin, Charles 1985. *The Correspondence of Charles Darwin*. Cambridge, Cambridge University Press.
Darwin, Charles (1987) *Charles Darwin's Notebooks, 1836–1844.* Eds. Paul H. Barrett– Peter J. Gautrey – Sandra Herbert – David Kohn – Sydney Smith. Ithaca/NY, Cornell University Press.
Davies, Brian – Michael Ruse 2021. *Taking God Seriously: Two Different Voices*. Cambridge, Cambridge University Press. https://doi.org/10.1017/9781108867375
Dawkins, Richard 1983. Universal Darwinism. In D. S. Bendall (Ed.) *Evolution from Molecules to Men*. Cambridge, Cambridge University Press. 403–25. https://doi.org/10.1017/cbo9780511730191.035
Dawkins, Richard 1986. *The Blind Watchmaker*. New York, Norton.
Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting.* Cambridge/MA, M.I.T. Press. https://doi.org/10.7551/mitpress/10470.001.0001
Dijksterhuis, Eduard J. 1961. *The Mechanization of the World Picture*. Oxford, Oxford University Press.
Dostoevsky, Fyodor 2003. *The Brothers Karamazov*. London, Penguin.
Dupré, John 2012. *Processes of Life: Essays in the Philosophy of Biology*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199691982.001.0001
Ellis, George 2011. Does the Multiverse Really Exist? *Scientific American*. 305/2. 38–43. https://doi.org/10.1038/scientificamerican0811-38
Fodor, Jerry 2007. Why Pigs Don't Have Wings: The Case against Natural Selection. *London Review of Books*. 29/20.
Fodor, Jerry – Massimo Piattelli-Palmarini 2010. *What Darwin Got Wrong.* New York, Farrar, Straus, and Giroux.
Friederich, Simon 2001. Fine-Tuning. E. N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2021/entries/fine-tuning/
Ghiselin, Michael 1969. *The Triumph of the Darwinian Method*. Berkeley, University of California Press.
Gibson, Abraham 2013. Edward O. Wilson and the Organicist Tradition. *Journal of the History of Biology*. 46. 599–630. https://doi.org/10.1007/s10739-012-9347-3
Grant, Peter R. – B. Rosemary Grant 2014. *40 Years of Evolution: Darwin's Finches on Daphne Major Island*. Princeton/NJ, Princeton University Press. https://doi.org/10.1515/9781400851300
Gray, Asa 1874. Scientific Worthies. *Nature*. 10. 79–81. https://doi.org/10.1038/010079a0
Harris, Sam 2004. *The End of Faith: Religion, Terror, and the Future of Reason*. New York, Free Press.
Hume, David 1779/1990. *Dialogues Concerning Natural Religion*. Ed. M. Bell. London, Penguin.
Kant, Immanuel 1790/2000. *Critique of the Power of Judgment*. Ed. P. Guyer. Cambridge, Cambridge University Press.
Knight, David 1990. Romanticism and the Sciences. In Andrew Cunningham – Nicholas Jardine (Eds.) *Romanticism and the Sciences*. Cambridge, Cambridge University Press. 13–24.
Laland, Kevin – Tobias Uller – Marc Feldman – Kim Sterelny – Gerd B. Müller – Armin Moczek – Eva Jablonka – John Odling-Smee – Gregory A. Wray – Hopi E. Hoekstra – Douglas J. Futuyma – Richard E. Lenski – Trudy F. C. Mackay – Dolph Schluter – Joan E. Strassmann 2014. Does Evolutionary Theory Need a Rethink? *Nature*. 514. https://doi.org/10.1038/514161a

Laland, Kevin N. – Tobias Uller – Marcus W. Feldman – Kim Sterelny – Gerd B. Müller – Armin Moczek – Eva Jablonka – John Odling-Smee 2015. The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions. *Proceedings of the Royal Society B*. 282. https://doi.org/10.1098/rspb.2015.1019

Lennox, James G. 1993. Darwin *Was* a Teleologist. *Biology and Philosophy*. 8. 409–421. https://doi.org/10.1007/bf00857687

Luther, Martin 1970. *Three Treatises*. Minneapolis, Fortress Press.

MacArthur, Robert H. – Edward O. Wilson. 1967. *The Theory of Island Biogeography*. Princeton/NJ, Princeton University Press. https://doi.org/10.1515/9781400881376

Morgan, Sue R. 1990. Schelling and the Origins of his Naturphilosophie. In Andrew Cunningham – Nicholas Jardine (Ed.) *Romanticism and the Sciences*. Cambridge, Cambridge University Press. 25–37.

Nagel, Thomas 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False*. New York, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199919758.001.0001

Newman, John H. 1973. *The Letters and Diaries of John Henry Newman, XXV*. Eds. C. S. Dessain – T. Gornall. Oxford, Clarendon Press.

Owen, Richard 1849. *On the Nature of Limbs*. London, Voorst.

Paley, William 1802/1819. *Natural Theology (Collected Works IV)*. London, Rivington.

Richards, Robert J. 1987. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago, University of Chicago Press. https://doi.org/10.7208/chicago/9780226149516.001.0001

Richards, Robert J. 2003. *The Romantic Conception of Life: Science and Philosophy in the Age of Goethe*. Chicago, University of Chicago Press. https://doi.org/10.7208/chicago/9780226712185.001.0001

Ruse, Michael 1977. William Whewell and the Argument from Design. *Monist*. 60. 244–268. https://doi.org/10.5840/monist19776022

Ruse, Michael 1996. *Monad to Man: The Concept of Progress in Evolutionary Biology*. Cambridge/MA, Harvard University Press. https://doi.org/10.4159/9780674042995

Ruse, Michael 1999. *The Darwinian Revolution: Science Red in Tooth and Claw* (2nd ed.). Chicago, University of Chicago Press. https://doi.org/10.1017/9781108672047

Ruse, Michael 2001. *Can a Darwinian be a Christian? The Relationship between Science and Religion*. Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9780511803079

Ruse, Michael 2003. *Darwin and Design: Does Evolution have a Purpose?* Cambridge/MA, Harvard University Press. https://doi.org/10.2307/j.ctv1qdqzvc

Ruse, Michael 2017. *On Purpose*. Princeton/NJ, Princeton University Press. https://doi.org/10.2307/j.ctvc773jn

Ruse, Michael 2021. *A Philosopher Looks at Human Beings*. Cambridge, Cambridge University Press. https://doi.org/10.1017/9781108907057

Ruse, Michael – E. O. Wilson. 1986. Moral Philosophy as Applied Science. *Philosophy*. 61. 173–192. https://doi.org/10.1017/s0031819100021057

Schelling, Friedrich W. J. 1803/1988. *Ideas for a Philosophy of Nature – as Introduction to the Study of this Science 1797 (second edition 1803)*. Trans. E. E. Harris – P. Heath. Cambridge, Cambridge University Press.

Schelling, Friedrich W. J. 1833/1994. *On the History of Modern Philosophy*. Trans. A. Bowie. Cambridge, Cambridge University Press.

Sedley, David 2008. *Creationism and its Critics in Antiquity*. Berkeley, University of California Press. https://doi.org/10.1525/california/9780520253643.001.0001

Spencer, Herbert 1852a. A Theory of Population, Deduced from the General Law of Animal Fertility. *Westminster Review* 1. 468–501.

Spencer, Herbert 1852b/1868. The Development Hypothesis. Reprinted in his *Essays: Scientific, Political and Speculative*. London, Williams and Norgate. 377–383.
Spencer, Herbert 1857. Progress: Its Law and Cause. *Westminster Review* LXVII. 244–267.
Spencer, Herbert 1860. The Social Organism. *Westminster Review* LXXIII. 90–121.
Spencer, Herbert 1862. *First Principles*. London, Williams and Norgate.
Weinberg, Steven 1999. A Designer Universe? *New York Review of Books*. 46/16. 46–48.
Whewell, William 1833. *Astronomy and General Physics (Bridgewater Treatise 3)*. London.
Whewell, William 1837. *The History of the Inductive Sciences (3 vols)*. London. Parker.
Whewell, William 1853/2001. *Of the Plurality of Worlds. A Facsimile of the First Edition of 1853: Plus Previously Unpublished Material Excised by the Author Just Before the Book Went to Press; and Whewell's Dialogue Rebutting His Critics, Reprinted from the Second Edition*. Ed. M. Ruse. Chicago, University of Chicago Press.
Whitehead, Alfred N. 1926. *Science and the Modern World*. Cambridge, Cambridge University Press.
Whitehead, Alfred N. 1929/1978. *Process and Reality: An Essay in Cosmology*. New York, Free Press.
Whitehead, Alfred N. 1938. Nature Alive. In his *Modes of Thought*. 202–232. New York, Macmillan.
Wilson, David S. – Edward O. Wilson. 2007. Rethinking the Theoretical Foundation of Sociobiology. *Quarterly Review of Biology*. 82. 327–348. https://doi.org/10.1086/522809
Wilson, Edward O. 1975. *Sociobiology: The New Synthesis*. Cambridge/MA, Harvard University Press.
Wilson, Edward O. 1992. *The Diversity of Life*. Cambridge/MA, Harvard University Press.

Gergely Kertész

On the Status of Teleological Discourse

A Confusing Fiction or a Description of Reality?*

"*Teleology is like a mistress to a biologist: he cannot live without her but he's unwilling to be seen with her in public.*"
J. B. S. Haldane

I. INTRODUCTION

There is a widely accepted view both in science and philosophy according to which teleological language is mainly a source of confusion and error as a description of nature. The view is held by many cognitive scientists (Kelemen 1999, see: De Smedt and DeCruz 2020) and also in much of metaphysics where teleological discourse is often depicted as the folk's way of systematically misrepresenting reality (Hartmann 1951, Rose and Schaffer 2017) and which is probably an unfortunate source of creationist intuitions at the same time (Kelemen 2004). In much of biology teleological language was viewed with suspicion from early on as it seemed to be a reminiscent of a misleading, pre-Darwinian way of understanding nature, therefore many biologists argued that this language should be replaced with proper evolutionary descriptions. To amend the situation, especially in the philosophy of biology a form of evolutionary teleonaturalism became popular, reinterpreting the teleological language of functions with direct reference to natural selection (Millikan 1984, Garson 2019).

This paper defends a different approach that allows for taking a large chunk of teleological discourse as veridical in a way that could be reconciled with naturalism. There are plausible theories on the table, according to which teleological statements can be systematically connected to the presence of certain type of complex physical systems, therefore teleological language could preserve its referential status by means of some form of reductive identification. This approach

* At the time of writing this paper worked as a researcher at the ELKH-BTK. The research presented in this paper was supported by Teloi.org and the Values and Science 'Lendület' research group at the ELKH-BTK. The author wishes to thank Daniel Kodaj for inspiring conversations while working on this paper.

doesn't deny that the application of teleological discourse might be misleading in many cases concerning the nature of things, but it holds that it is definitely not as guilty as charged and has a fairly good reason to be.

The basic strategy for defending teleological discourse runs as follows: teleological intuitions and teleological discourse are a product of evolution and plausibly it has evolved to track the behaviour and presence of complex self-maintaining systems (basically organisms) in nature, the workings of which might involve tool use and also certain social structures or superorganisms. Such systems have an internal organization that makes sense why humans explain their activities in teleological terms. This view does not deny that humans might use teleological language to describe systems where its use is not justified, that they might overuse this tool, but as the cost of applying it in a too permissive way, the cost of erring on the safe side, isn't high we shouldn't be surprized about that. Naturally, starting from wrong assumptions concerning e.g. the broader context, teleological descriptions might be applied to systems that do not serve a purpose, do not have a function, however, such mistakes can be corrected. The suggestion will be that instead of concentrating on misfiring heuristic applications of teleological discourse, it would be more beneficial to treat teleology as a property of certain types of complex systems, reducible to some system level properties similarly to other macro properties such as temperature or mechanical hardness and other practically useful macro-physical properties.

The gist of the idea presented here is this: if some version of the reductive identification of teleological systems is attainable, we should treat teleological discourse as tracking real distinctions in nature. To be able to run the main train of thought in this paper the theoretical possibility of such identification would be sufficient in itself, but I will also suggest that it is more than a possibility, it is rational to think that it is a plausible option. It is widely known that such attempts were already pursued by early cybernetics from the middle of the 20th century. Some philosophers and biologists tried to explain the apparent goal-directedness of certain systems on grounds of internal features like feedback-based organization (e.g. Braithwaite 1953, Sommerhoff 1969). In the beginning this attempt was also endorsed by mainstream biologists like Mayr (1974) who introduced the notion of 'teleonomy' to describe the apparent purposiveness of the living in need of explanation, but by the 1980's this project started by cybernetics was considered to be largely unsuccessful (see: Bedau 1992), largely because the normativity of teleological language seemed to be unexplainable by the theoretical means suggested. At the same time a somewhat different strand of theoretical biology, general systems theory created a holdout for analysing the organizational features of organic life introducing the concept of autopoetic systems (Maturana and Varela 1980). As I will explain below, a continuation of this tradition is what provides credibility, plausibility for reductive identification.

From the point of view of this tradition it is not implausible to think that teleological systems are real in the sense that they can be identified and differentiated from other types of systems based on their organizational features independently of folk teleological attribution. This provides a good enough basis for thinking that teleology is something that is reducible in an ontologically conservative manner, meaning that the higher-level property, as that construct does not turn out to be problematic, too imprecise or empty, is not eliminated by the reduction, but is conserved by its identification with some base properties (Savitt 1974), similarly to e.g., mechanical hardness (see Gilman 2009). In the cases of most macro-level physical properties micro-level reduction is not considered to be the elimination of the macro property even if substitution is made possible by the identification, instead it is considered to be matched onto more fundamental entities, properties and their configurations. Similarly, I will argue, that it is plausible to think that teleology is not an ontologically fundamental property, but it is a property of certain type of complexes organised from simpler elements.

II. THE ORGANIZATIONAL VIEW OF ORGANISMS – HISTORY SKETCH AND OUTLINE

According to a more and more influential new theory in the philosophy of biology, that is rooted in the tradition of the already mentioned general systems theory, to be a living entity is to be a self-determining system. In this section, I introduce this approach in some detail starting from a historical perspective (for a deeper discussion see Moreno and Mossio 2015, Mossio and Bich 2017).

Some broader context first. One might ask what brought the organism to the fore in recent biological theorizing? For a long time, the focus in the philosophy of biology was on conceptual issues surrounding evolutionary theory and some of its consequences. In the neo-Darwinian synthesis, the organism was reduced to the genes, they became the real agents of evolution (Hamilton 1964), a view made popular by Dawkins (1976). Organisms in this perspective were only the 'vehicles' or the 'interactors' of the real replicators, the genes. In this orthodox theory the focus was on two levels of analysis: on genes and on populations of genes. Organisms were omitted for convenience's sake. More recently, this trend started to dissolve, and the organism is having its renaissance. Walsh (2015) highlights that developmental biology, new research on ontogenetic developmental processes, a special interest in epigenesis, evo-devo theories and in the effects of niche construction created the need for a revaluation of the role of the concept of an organism in biology. And it is true that the models of the new evolutionary synthesis under construction, recently getting represented even in university level textbooks, are considering not only the mentioned two levels,

in the new systematization the level of the organism became a level of analysis on its own right.

It is quite probable that this development is less surprizing for people interested in 20th century general systems theory originated by Bertalanffy (1951, 1968). That tradition had at least two strong arguments for highlighting the role of organisms from the outset. The first reiterates an old idea (1): prolification presupposes self-replication in time (Csányi 1982), or in other words, the evolutionary process presupposes the existence of self-maintaining systems. Replication in space, prolification can only appear if self-replication in time is already in place. Staying alive, maintaining organizational invariance in the face of constantly shifting environmental conditions, is the problem that has to be solved first before reproduction can become an issue at all. The second reason (2), the one that came to prominence more recently (McLaughlin 2001) is that evolutionary theories of biological function that identify function based on past selection history have a hard time with explaining the function of a biological trait created anew.

The other prominent teleonaturalist theory of biological functions, first advanced by Millikan (1984), bases everything on natural selection claiming that something has a function at present because it had a certain history, the function of a phenotypic trait we observe is what it was selected for. One important challenge to this view was the presence of vestiges, like the appendix. Suppose that it is true to say that it used to harbour gut bacteria, but it lost that function and at present it serves no function. The presence of the organ can still be explained on grounds of selection history, but it would be absurd to say that it has the function it was selected for in a bygone age. There are possible fixes to this problem. One could say that we should focus on the recent, or more immediate past of a trait (Godfrey-Smith 1994) and check whether it contributed to fitness in the period in question (Schwartz 1999).

However, even if those fixes work, we can also say that a functional description highlights a causal contribution to the workings of a particular living system at present whatever its history was. It might have a good pedigree in terms of its history, but what decides its fate and role is what is taking place at present. This is what gets highlighted by the already mentioned case of newly invented traits. This issue is usually discussed based on Davidson's well-known Swampman thought experiment. The Swampman is instantaneously created in some swamp through an improbable cosmic coincidence of quantum events, but still, it has the very same biological features as any human being. However, as it is not part of a lineage and therefore lacks a selection history, its organs cannot serve evolutionarily established functions. One might object, as some did, that this thought experiment is empirically highly implausible. To that I would answer, the example still clarifies the theoretical difficulty nicely and it describes a scenario that is quite close to the case of the appearance of new variations or

mutations in the course of evolution. A useful, but evolutionarily new invention resulting from recombination or mutation functions beautifully in the talented young organism, even though it lacks any kind of selection history.

As I already said, the self-maintaining organisation of the individual organism is a prerequisite for selection processes, but it is also for the attribution of functions: a functioning part of a living system is good for that organism, however a dysfunctional part is bad, so purpose and function imply normativity at the level of the individual in a way that is not implied by other properties in nature. Which means that the ultimate ground for function ascription should belong to individual organisms themselves independently of their histories. However, histories are definitely not dispensable. Evolutionary explanations are important in themselves, but evolutionary histories don't exhaust the bases for function attribution and are not even the most important reference point for it.

A final note that might be surprizing for some readers. As Michael Ruse shows in his paper in this very volume, Darwin himself could be called to defend the view that organisms have some form of immanent teleology and it is exactly the origin of this adaptedness and adaptivity in organisms that gets explained by natural selection. A distinction should be made between Platonic teleology, where the source of telos can only be a Creator and Aristotelian teleology, where it is immanent to the being that has the telos. In the second case the function of the traits and behaviour belongs to the organism itself, not to its history, not to its maker, and the explanation of its presence is an altogether different issue (for a more detailed discussion of this distinction and its uses see Ariew 2007).

The *organizational view of organisms*, versions of which were already advanced by 20th century systems theorists, has a surprisingly long history. It is older than Darwinism. Let me give the reader a sketchy outline of that history. In the Aristotelian tradition organisms were defined by reference to features like self-motion, autonomous functioning, and separation from the environment (see Gelber 2021). These concepts describe organisms in terms of observable behavioral patterns contrasting them to purely physical entities. By focusing on the more general capacity of self-maintenance, one approaches organisms in terms of their distinctive internal organization that sets them apart from other types of complex physical systems. This more modern concept also has a prehistory in philosophy, most notably in Kant's work on purposiveness (cf. Moreno and Mossio 2015. xxiii-xxv). Kant held the view that only our limited cognitive capacities make us interpret living things as purposeful. However, at the same time he admitted that the reproductive and regenerative capacities of the living were inexplicable by the means of the physical science of his age. As a resolution to this tension, he coined the term 'self-organization' and described organisms as naturally purposive, characterized by a kind of immanent teleology, meaning that their internal mechanisms serve the purpose of maintaining the whole. However, we should note, that as he could not reconcile this picture with the science

of his age, as he could not accommodate circular causation with the physics that was available, therefore for him teleology worked only as a regulative principle of reason and teleological descriptions of nature were considered to be ontologically non-consequential. For the modern view introduced below, circular or recursive causation creates no such puzzle, so self-organization features can be directly connected to lower-level dynamics.

Most contemporary conceptualizations of the living and of organisms in the philosophy of biology connect back to Kant's work on natural purposiveness via the second half of 20th century tradition of general systems theory (e.g. Bertalanffy 1968, Maturana and Varela 1980, Kampis 1991). Living systems are understood to be systems that self-maintain or self-replicate over time where this feature defines the fundamental goal of their activities. The parts of such systems actively contribute to the regeneration, recreation of other parts of the system. This creates a closed network of regenerative, functional connections between the different kinds of parts that, as I will show below based on work done by mainly former students of Maturana, amounts to a defining feature of these systems.

The organizational view introduced in this paper is a continuation of the systems theory tradition. To make the gist of it more intuitive, let us start with the idea of minimal self-maintenance, the proper understanding of which brings us closer to a definition of the kind of self-maintenance that defines organisms. All self-maintaining systems, including non-living ones, contribute to the maintenance of their own conditions of existence (see Mossio and Bich 2017). A candle flame self-maintains in the sense that the flame persists via maintaining a cycle: the heat it radiates by burning the vax melts the remaining wax that provides further fuel for radiating heat. At the same time hot combustion products are carried upwards, which creates a constant influx of oxygen rich air from the sides, also contributing to flame-persistence till the point when the vax runs out. What we observe as stability in such systems is a result of this cyclic flow. We all know that this system is fragile and the flame disappears swiftly without an external influx of energy. Candle flames are not in a stable internal state, like an atom sitting in a potential well, but in an instable, relatively high entropy state and exactly because of that their persistence hangs on running that cycle.

Candle flames are simple self-maintaining systems, which means that they are undifferentiated. They have no real parts, meaning that there is no internal division of causal labour inside. Any arbitrarily chosen proper part of the flame does the same kind of work, they melt and burn the vax. By contrast, living systems have functionally differentiated parts organized into a causal division of labour, each contributing differently to the maintenance of the whole (see Mossio and Bich 2017). All parts of such systems realize functions that serve the fundamental goal of self-maintenance at the level of the whole. When it comes to such systems the attribution of functions can be based solely on the identi-

fication of the role a part plays in self-maintenance, which also means that biological functions can be defined in an interest-independent manner and without reliance on the evolutionary history of organisms. This approach to function was first systematized by McLaughlin (2001).

Let us take a look at a section of such cycles. The repeated contractions of the heart in animals maintains blood flow and therefore oxygen and nutrient levels throughout the body, contributing to the persistence of the organism via maintaining other organs that, in turn, contribute to the maintenance of the heart. All parts of such systems 'work' for their continued existence by maintaining the right internal conditions and the influx of energy and building blocks for other parts and thereby for the whole. As two classic authors of systems theory put it (Maturana and Valera 1992), "being is doing" for self-maintaining systems that exist in far from thermodynamic equilibrium states. Without the constant regeneration cycle going on things would swiftly degrade.

From early on organisational views built heavily on the notion of a *boundary condition* (sometimes also called a *constraint*), a causal notion that is required to make real sense of the complex causal division of labour in organisms and with that of the notion of self-determination. Here I will explain this concept based on examples and only at a more intuitive level as it is not the main focus of the present paper.

The essence of the organizational view is that living systems are characterized by a circular causal regime that reproduces the *internal boundary conditions*[1] necessary for various processes in the causal cycle itself. But before things get conceptually too complicated let us see what is a boundary condition more generally and how is it different from an ordinary cause, a causal factor? In one sense it is just a causal factor, but it has some important features not considered by everyday causal talk, neither by received theories of causation (like interventionist theories (Woodward 2005), the INUS account (Mackie 1974), or process views (Dowe 2000)), that are important for physical explanations and calculations in the physical sciences. Take the case of blood circulation again. Its function is to deliver nutrients, oxygen and hormones to all parts of the body. The vein walls are indispensable constrains that channel our blood to its destinations. In the parlance of the organizational view, their presence is the most important *boundary condition* for the process of blood circulation to reach its end even if not the only one. However, in simple causal parlance their role cannot be differentiated from other *causal conditions* or causes of blood circulation, like the pumping of the heart. Blood circulation obviously has a lot of further causes, causal factors like external atmospheric pressure, internal body temperature range, etc., many

[1] By *internal* it is simply meant that the boundary condition in question is part of the causal cycle as it is created or regenerated by it, and it also regenerates or creates another boundary condition in the cycle. Many causal cycles are not like this. More on this later.

of which are not only causes, but also boundary conditions for circulation at the same time.

So how can we differentiate boundary conditions and other causal conditions in general? The best way is to think of boundary conditions as constraints the constant presence of which is required for some other process to be able to run its course. A constraint is such that it determines, shapes some dynamics, but the dynamics leaves it largely intact. This might be familiar from pure mechanics: the surface of the table constrains the movement of a ball rolling on it. The same initial momentum leads to different trajectories depending on the exact shape of the surface as a boundary condition. One can connect this concept back to causation by distinguishing between stable/constant and unstable/variable causal conditions. In most examples of causal processes, such as a house fire, many of the causal preconditions are used up. The fire consumes oxygen and combustibles leaving ash, cinder and CO_2 behind. By contrast, atmospheric pressure or the absence of firefighters are causal conditions that need to be constant for the combustion process to go through. As these examples show, the notion of a boundary condition is not inherently teleological. The same notion applies in the context of pure mechanics as in the biological examples discussed.

For some biological process to run its course most of its causal conditions need to remain constant while the process plays out. What makes the situation peculiar is that many of these constant conditions are the products of the organism itself. For example, protein synthesis requires both amino acids and enzymes as catalysts, with the former being transformed or used up in the process while the latter remains a constant throughout the process. The presence of amino acids and the presence of enzymes are both causal conditions of protein synthesis, but the enzymes are not used up by protein synthesis; rather, they channel the process toward a specific outcome. In turn, enzymes themselves are products of other processes of the self-maintaining cycle of the organism. This is what makes them internal boundary conditions of the cycle. they are both created by the cycle and are also indispensable enablers of further steps in the cycle. Unlike e.g., external atmospheric pressure.

Now we have all the required conceptual resources to formulate what is peculiar about biological systems and about biological self-maintenance. Any process that is carried out by an organism requires a host of boundary conditions. Many of them, like the enzymes, are internal to the system and internally produced by it and while naturally degrading they are always reproduced by other parts the system itself to serve their function as a constraint again, leading to a circular causal regime where the internal boundary conditions in the system are the causal conditions of each other's reproduction processes within the self-maintenance cycle. An enzyme makes possible a certain kind of protein synthesis process, but that very enzyme is synthetized with the patronage of another enzyme and so on. In general, the organizational view states that all internal boundary

conditions in a living system are such that they are produced with patronage of some other internal boundary conditions in the system and the system is closed for this relation. Biological functions are nothing more than the roles of the internal boundary conditions in reproducing other boundary conditions.

After setting the stage I can also introduce the notion of self-determination. An organism determines its own fate, persistence into the future, by creating many of the boundary conditions, at least the internal ones, and built on that, in most organisms, even some of the external boundary conditions, circumstances that allow it to exist further. We are not talking about a lucky cycle for which the circumstances are just right accidentally, a self-determining cycle creates the conditions that allow it to persist, justifying the special name for the kind.

To sum up, all internal boundary conditions (BC) in an organism are produced by and within the system itself and this is what basically makes it into a self-determining system. This what Moreno and Mossio (2015) call the '*Closure of Constraints*':

- Every BC_i in the system is a BC for the regeneration of at least one other BC_j in the system
- Every BC_j in the system is subject to at least one other BC_i in the system

What a BC_1 does contributes to the existence of BC_1 itself and the living system itself by its contribution to the existence of other constraints BC_2, ..., BC_n. Note that such closure does not entail that there are no further external conditions for the existence of an organism or that the organism does not have some effect on its own external boundary conditions. The cycle itself runs only in the presence of certain external conditions (like e.g., gravity, atmosphere, etc.). However, this closure of boundary conditions in the self-maintenance cycle of living systems is what makes them special in terms of their internal relational organization according to the organizational view.

From this it should be clear that having a closure of boundary conditions is more than running a circular causal regime, that might also occur in inanimate nature. Consider the hydrological cycle: water evaporates from open waters, it forms clouds, then precipitation occurs, then rainwater follows the slope of the land and flows back into the oceans. The slope is a boundary condition of the cycle, but it is not caused by other processes that are part of the cycle. The evaporation, rainfall etc. are not boundary conditions for the slope. A cycle like the hydrological has no internal boundary conditions.

A more formal definition of organizational or biological closure can be created based on these ideas that does justice to the view of organisms as "unimaginably complex self-maintaining storm of atoms [that] moves across the surface of the world, drawing swirls and clots of atoms into it and expelling others, always maintaining its overall structure" (van Inwagen 1990: 87). *Organizational Closure* (def.):

> Some simples (e.g. physical building blocks), the *x*s, compose an organism if:
> The *x*s can be grouped into the x_1s, x_2s... x_ns such that (i) for any i ($1 \leq i \leq n$) there is a j such that the activity of the x_is is a boundary condition of the activity of the x_js, and (ii) for every i, the activity of the x_is is a boundary condition of the collective activity of the *x*s (that is, of the causal cycle as a whole). (for a more detailed discussion of the definition see: Kertész & Kodaj 2023)

The above definition allows for the fact that organisms change (1) their building blocks (*x*s) constantly and (2) their mode of self-determination in response to the environment, switching from one self-determining regime to another, changing behaviour, modes of feeding, digestion etc., all the while maintaining the internal cross-dependence defined by (i) and (ii). So, we have a definition that is a good start for more informative definitions of organisms, or living systems and should be enough for the purposes of this paper.

III. REFERENCE FOR TELOS AND FUNCTION ASCRIPTIONS

In this paper, I am not committing myself with respect to the validity of the organizational view. Even though I find it to be a promising research program, I only use it as the best available theory of organisms that attempts to define them as a type of complex physical system in contrast to other type of physical systems. I don't claim that the criteria developed by its proponents are correct, sufficient or easily justifiable in an empirical sense. Even though its proponents have already created more formal systematizations of this idea (Mossio et al. 2009) that bring it closer to computer simulation based tests and other more practical applications, at the moment this is first and foremost a promising theoretical construct. What I accept without critical discussion here, is that the criteria provide at least a necessary condition for the identification of organisms and maybe more than that with somewhat blurred boundaries of identification. Therefore, what I am interested in is the following question: if the organizational view were a successful theory of organisms what would be the consequence of that for the status of teleological discourse?

The short answer to this question is that it could provide a firm basis for the idea I proposed in the introduction. If we take self-determining systems and their parts to be the entities referred to by teleological statements, a lot about the uses of teleological discourse gets explained in a reductive sense and, on the other hand, we can also shed light on what happens in erroneous, or promiscuous applications and how could we correct those mistakes.

Let us sharpen our understanding of the referential bases of teleological, functional language. First, we said two things: (a) if the organizational view is true, then organisms are self-determining systems. (b) teleological attributions

have the function of tracking the presence and explaining the behaviour of organisms, that are, according to (a), self-determining systems. In analysing the connections between teleological and systems language in more detail, I follow Moreno and Mossio (2015). First, what makes it justified to attribute purpose, a telos to organisms or their behaviour? Well, it is certainly not that they or their parts have minds or intentionality. Here, I need to make an important distinction. Intentions and intentionality only appear on the scene with the presence of a mind, cognition and representations and here I will use these terms in a way that respects this understanding (in contrast to e.g. Daniel Dennett). Having a telos is simpler, more basic feature than having a mind. When we search for the end of e.g. some anatomical feature, we certainly don't attribute a mind or intentionality to that feature. We try to situate it in a system and especially when that system is itself a mindless creature than the only plausible question is this: how does it help the creature to live, to stay alive? In the parlance of most system theorists the question is, how does it contribute to the self-maintenance of the organism?

This latter question is key for two reasons. First, it shows that biological teleology is independent of the attribution of mind and intentionality. Second, it makes it clear that something can only serve a function if it is situated in the right kind of context. The only approach that makes the attribution of function both objective and independent of the particular interests of an observer is based on the organizational view of organisms. In that perspective, the function of a trait (if it has or had any), e.g. an anatomical feature, can be identified by locating its causal contribution to the self-maintenance of a self-determining system it is (was) a part of. That is the right 'context' in which function can be attributed. Any activity/trait of a self-determining system has intrinsic relevance to itself, to its existence to the extent that its persistence depends on the contributions of those constraints that the activity in question maintains in the system. However, if something is (was) not part of such a system any attribution of function is only in the eye of the observer. So, self-maintenance can be identified as the fundamental telos of a living system to which all functions of its parts are subordinated. Notice that a function doesn't belong to a part, it is not intrinsic to the part. The same gene or neurotransmitter might serve different functions in different species. A function is a relational property, it identifies the role of a part in a self-determining system and the system as such cannot change into another system without a change in at least the contribution of some parts to self-maintenance and if a function changes the systems changes. So, the functions of the parts are intrinsic to the system in the same way as its basic telos.

Functions are also supposed to explain the existence of function bearers. E.g. the heart's activity of pumping blood explains its presence and persistence. Notice that self-determination allows such explanations to make good sense. The

heart does contribute causally to its own persistence via the self-determination cycle of the whole system it is a part of. For the same reason, self-maintenance serves as the basis for the normativity of functions and teleology. We expect organic parts to function in certain ways, exactly because they stay present only if they do their work in the self-maintenance cycle. Interestingly, biology textbooks, even books on physiology are full of normative descriptions like: 'in order to', 'demand', 'need' and so on. There is a lot of discussion of control systems in biology. For example, in genetics textbooks they tell us that a cell can 'control' the proteins it makes by 'controlling' gene transcription. This language implies distinguishable states of a system from which some are preferred over others. In a self-determining system that is a meaningful evaluation if we take self-maintenance as the basic telos into consideration. So, 'control' is a teleological and normative notion, it is done for the sake of self-maintenance. It can be successful and it can fail from the internal perspective of self-maintenance. Biologists tend handle such language with distancing gestures and cautionary remarks about language use. However, in my view, when we accept that teleological idioms refer to self-determining systems and their parts, we accept that teleological language is basically innocent, requires no distancing gestures, as it has a respectable reference base in the physical realm.

IV. REDUCING OR ELIMINATING TELEOLOGY?

On grounds of the above I suggest that teleology and function are properties that are reducible in a similar sense as temperature or mechanical hardness is reducible. Maybe because of multiple realizability considerations identifications can only be created locally (see Kim 1992), but I suspect that the analogy with known cases of physical reduction are stronger than one would think. Let me start by introducing the classic case of reduction based on Nagel's conceptualisation. Starting from there I will show that organisational properties like being a self-determining system are sufficiently similar to aggregate properties like mean kinetic energy of gas molecules, or calculations based on the strength of chemical bonds in solid matter.

Usually Nagel's theory of theory reduction (Nagel 1961, 1970) is taken to be a theory of reductive explanation, not of ontological reduction and this might create some confusions, so let me shed some light on this issue. His bridge-laws do serve explanatory purposes because their function is to connect different theories using different descriptions of nature, allowing the derivation, the explanation of the reduced theory or terms of the reduced theory by the reducing theory. It is true that the possibility of derivations like that does not imply the necessity of full-blown inter-level identification. However, according to Nagel, in most cases bridge-laws also declare the identity of the properties, co-refer-

ence of the different terms or term constructs of the higher and lower-level theories (see: Fazekas 2009. 305–306).

The qualitative distinctness between the properties talked about in the two different theories is what makes reduction an interesting achievement. When temperature in a volume of gas is reduced to the mean kinetic energy of the gas molecules in the volume in question, the bridge-law connecting the two shows "that what are prima facie indisputably different traits of things are really identical" (Nagel 1961. 340). The two prima facie different terms refer to the very same thing. The qualitative distinctness is quite straightforward here: no lower-level gas molecule has a temperature, that term is meaningless in the realm of molecules. Those particles have a few basic properties like space and time location, charge and kinetic energy, but only the last is relevant for the reduction. In statistical thermodynamics the temperature of a volume of gas equals to the mean kinetic energy of the ensemble of particles that make up the volume of gas. The reduction achieved imply that the terms temperature and mean kinetic energy of molecules in the volume refer to the very same thing under different descriptions.

So, it is useful to differentiate two aspects of reduction (see: Crane 2001). First, explanatory reduction, when what is explained by a higher-level science also gets explained based on a lower-level science. Explanation expresses an asymmetric relation. The lower-level science explains a term of the higher-level science, but not the other way around. Explanatory reduction does not require inter-level identity between the entities assumed to exist, temperature could be reduced to a different theoretical construct in solid matter and in gases and according to some theorists this is the case (see: Sklar 2015).

Secondly, we can talk about ontological reduction, when it is shown that a term in the higher-level theory refers to the same entity that the lower-level reducing theory is talking about, just under a different name or complex description. This relation is symmetric. The two descriptions are ontologically reduced to each other as the terms co-refer. The identification of two entities does not eliminate either. Claiming that Charles Bronson is Charles Dennis Buchinsky does not imply that either is non-existent. So, although reductionists like to use the phrase that something is "nothing over and above" or "just is" this or that, reduction via bridge-laws is not the same as elimination.

Identity reduces the number of autonomous entities we should accept into our basic ontology but allows that there are different scientific categories in sciences at different levels that pick out the same real thing. In the case of temperature, we might say that only the particles, molecules with one of their basic properties are necessary to compute the temperature in the volume from lower-level information. This shows that the particles have ontological priority. However, we would not say that temperature is not a real property. It is real exactly because there is a procedure that show us how to connect it to more basic things

in nature. The bridge-law allows for a substitution of terms between different languages, but it does not imply that the term temperature is useless or mistaken, like the term phlogiston for a model of combustion. In what follows I will argue that a similar situation holds for teleological discourse as for temperature.

Let's start with the case of the statistical mechanical reduction of temperature. The bridge-law connects the macro- level property temperature of a volume of gas (T) and an aggregate property constructed from the micro-level property of kinetic energy that characterizes each particle in the same volume and it is defined as the mean kinetic energy of molecules in the volume (MKE), a property of an ensemble. So [T is MKE]. This is a very imprecise qualitative formulation, but here it is enough to say that MKE explains T in lower level-terms. But the statement also implies that T and MKE refer to the very same thing and if they co-refer both terms refer to the same thing in reality.

How could this work in the case teleology? First, we need to connect the property 'teleological' (TE) to the term we defined as *Organizational Closure* (OC). What we mean by having teleology gets connected to the complex system property of *Organizational Closure*. The complex system property explains what we mean by being a teleological system as an intrinsic property. So [TE is OC]. So far so good. The difference really comes out when we realize that TE is a quality and we cannot give it a quantitative interpretation. Unlike in the case of temperature, we can only say that the system *does things for its own good*, which is generally true of teleological systems, and then we have to start detailing its behavioural capacities in service of itself. The same goes for OC, which is only a general organizational feature, but to give any further qualification to it we would need to start to spell out the component level organizational features of the system, the functions that the different parts play in self-maintenance and to show that the cycle is really closed. Such characterization would be too detailed and overly complex for proper treatment in a paper like this. I could only give partial, surface-level examples from the life sciences that only highlighted functions of particular anatomical features. Notice that the only thing that is relevant for the discussion of the reduction proposed here is proving that a particular system is really an instance of OC. OC as such is definitely a multiply realizable feature, different organisms have fairly different self-maintenance cycles, which is displayed in their different behavioural patterns, homeostasis, organs, etc., but those differences are irrelevant with respect to being a teleological system. According to the proposed view, teleology as such consists in being a self-determining system that is an instance of OC. Nothing more, nothing less than that.

At first sight this might sound too different from the case of temperature. The difference can be understood if one looks at the general differences between the underlying systems (see Kampis 1991. 207). A volume of gas exists in the range of disorganized complexity. In such systems the individual degrees of freedom of the constituent parts do not play a direct role in the behaviour of

the whole, the parts are quite uniform, their properties can be easily averaged into some gross behaviour by simplifications along certain dimensions. In contrast, a biological system exists in the range called organized or inhomogeneous complexity where there are a wider variety of parts some of which contribute to the behaviour of the whole disproportionately. However, even though this difference is important for understanding the workings of these different kinds of systems more generally, this difference is no obstacle to the identification of the capacities of the organization as a whole and the component level organizational features of the system, the system of functional relationships between the parts (Fazekas–Kertész 2011, 2019). This is exactly what full-fledged 'mechanistic' explanations in the life sciences should ultimately aim for (Bich–Bechtel 2021) instead of just analysing such systems analogously to classical machines as much of the literature on mechanistic explanation in the life sciences in the last two decades did (e.g. Craver 2007).

V. ELIMINATIVISM, FICTIONALISM CONCERNING TELEOLOGY IN BIOLOGY

But before I delve into the discussion of the possibility of reduction more deeply, I should highlight that concerning teleology the eliminativist and the closely allied fictionalist attitudes rule supreme in biological theorizing and this is exactly what I would like to oppose in this paper. Although the context and the argumentation is different, in the philosophy of mind a parallel attitude became fashionable with the advent and development of neuroscience. In the view of the proponents of eliminative materialism (Chruchland 1981) as there are no mental states as depicted by folk psychology, both the identity theory and functionalism are trying to do something absurd, to reduce a non-existent to neural activity. This view presupposes that even though folk psychology is a theory of mind it is a useless, an outright wrong theory of the mind. Just as late 18th-century chemical theory did not try to save the concept of phlogiston in the context of molecular theory but simply dispensed with it and replaced it with oxygen theory, so the entire mentalistic vocabulary of folk psychology should be eliminated on behalf of the descriptions of advanced neuroscience. What I would like to point out below is that a similar approach to teleological language rests upon a mistaken attitude towards its uses in describing reality.

The eliminativist attitude in biological theorization has mostly to do with the dominance of evolutionary thinking and certain philosophical, metaphysical uses of Darwinism. The basic attitude is this: all apparent teleology was explained away by evolutionary theory, nothing remained and at the same time teleological language and explanation is unscientific so it should be dispensed with altogether. Darwin provided a mechanistic explanation for the changes ob-

served in the history of life and so beliefs in the purposefulness of historical change in nature are mistaken, what really takes place is a combination of blind variation and selective retention of the fortunate forms that are more fitted to the environment. This view became important for scientists and philosophers alike as an argument for a monist, naturalist worldview as it was the Darwinian perspective that provided the best argument against natural theology and for dispensing with the idea of a Creator in the context of understanding biological nature. Historically the architects of the modern synthesis of evolutionary thinking interpreted Darwin's role in the debate over the place of teleology in biology, as providing the theoretical tools for "getting rid of teleology and replacing it with a new way of thinking about adaptation" (as Michael Ghiselin claims in his preface to a modern edition of Darwin's work on orchids, see Lennox 1993) and thereby making a huge step towards an integrated naturalistic worldview.

This eliminativist stance that considers teleology to be a false relic of pre-Darwinian thinking is closely allied with a form of fictionalism according to which teleological descriptions should be seen as metaphorical and only serve as replaceable abbreviations, shorthands for proper evolutionary accounts. E.g. Maddrell (1998) describes what even professional biologists do regularly "for the sake of saving space" this way: "the proper but cumbersome way of describing change by evolutionary adaptation substituted by shorter overtly teleological statements". Ghiselin (1994) argued against those who found that Darwin can be rightfully interpreted as someone who saw himself explaining the origins of the immanent teleology in organisms claiming that Darwin's thinking is not teleological, only his language is, he only uses teleology as a metaphor, a kind of 'as if' description. As Michael Ruse reminds us in his article in this volume, in earlier work Ghiselin even went as far as to claim that when Darwin uses teleological language in his book on orchids he is doing "metaphysical satire".

Why do we need such fictionalist accounts of teleology? The answer is obvious, the extensive use of teleological language by both layman, but also by knowledgeable experts requires some form of explanation if we firmly believe that this language is a misrepresentation of biological reality or that it is misleading. We are allowed to say that some organ serves a function, has a purpose only if we can replace that language with some scientifically respectable mechanistic parlance. But ultimately function and purpose should be considered eliminable items from our dictionary of reality, as things only seem to have a purpose, but they don't have a purpose really. This is where I would like to suggest, that it might be better to consider the option of taking teleological language more seriously.

There is a lot to agree with concerning the intentions behind the eliminativist/fictionalist conceptualization of teleological language, but one should also see that what is true about the process of evolution doesn't necessarily apply to organisms themselves. If one takes the caution against teleological language

as a call to reject the idea that the adaptedness of organisms is a result of conscious design or some other intelligently guided process, etc., it is a fair point. But if it is about rejecting the idea that organisms in themselves, their parts and behaviour have purpose or function, the situation is much less straightforward. The reason why Mayr (1974) and some other prominent theoretical biologists favoured the introduction of the notion of teleonomy into the vocabulary of biology to replace teleology, the reason why they were interested in cybernetic accounts of apparently goal-directed behaviour (Sommerhof 1969) was exactly that the recognition that the properties of the organism should be handled separately from its history. I think it would be better to follow the path these theorists started to walk concerning teleology and that path leads to a system theoretic account of teleology.

Some contemporary Kantians provide a more constructive account of how and why teleological metaphors are useful (Breitenbach 2009). They argue that attributing teleology serves as a useful heuristic in the search for proper causal-mechanistic explanations of whatever organisms do. For a Kantian teleology can only serve as a regulative principle, our limited cognitive faculties are compelled to see organisms as purposeful, but teleology itself is what Kant calls a transcendental illusion. Therefore, for Kant, mechanistic science cannot explain teleology and therefore teleological descriptions have no ontological implications. At the same time, such descriptions create an analogy with purposeful human creation which provides a useful heuristic device for understanding how things really work, without committing the user to anything ontologically consequential.

This is a respectable view, which could also be categorized as a form of fictionalism simply because human planning and creation is not the real source of functionality in biological systems. However, one should also notice that Kant's inability to find a place in his system for the recursive kind of causation that characterizes organisms is problematic from the perspective of contemporary science. He saw them as causes of themselves, but such self-referential, recursive workings were incompatible with the linear view of causation that his idea of natural laws and scientific explanation implied. But we are not in Kant's position.

Contemporary science is well-equipped to handle both non-linearity and causal loops and this opens the door for taking teleological language seriously. Self-determining systems as described above are involved in non-linear dynamics and they are running energetically, thermodynamically open, but otherwise closed causal loops. Describing such systems as causes of themselves would be imprecise, but describing a cycle in self-maintenance as causing the next cycle and the functional parts of the system as causing the construction of a new token part of the same type that was instantiated before and thereby preserving a token of the type of organization that defines the organism in which the cycle is

running, is feasible. This is the perspective of general systems theory advanced most prominently today by Moreno and Mossio (2015). But then taking teleology only as a heuristic that helps us with projecting our own teleological activities onto mechanisms is an unnecessary restriction. A different angle becomes possible concerning the task: understanding organisms on their own right and by doing that probably understanding the reason why they attract teleological descriptions so readily in contrast to objects of inanimate nature. This highlights one reason why the comeback of the notion of an organism in the last decade is a quite significant change for theories of teleology.

VI. SAVING TELEOLOGICAL INTUITIONS AND LANGUAGE

To close the previous thread let's get back to eliminativism. Is there still a good reason for eliminating teleological language? Answering the question, I will take it that teleological language is akin to mental discourse, folk psychology, in the sense that it is a kind of theory, it provides a more or less fitting model of a section of reality. As a theory it can be useful, predictive or not as any model or theory. I think it is arguable that this discourse is relied upon both in everyday life and in science and the second kind of application is a continuation of the first, the same framework utilized more systematically, methodically. I also think that the cognitive module responsible for teleological intuitions and language is a product of evolution itself and it was selected for its usefulness in tracking the presence of and in understanding the behaviour of self-determining teleological systems from predators to prey, possible allies and enemies from other humans to poisonous bugs. Research in cognitive psychology portrays our teleological cognitive module as a core part of our cognitive toolkit (Csibra–Gergely 2007) pointing towards similar uses. This line of research shows that goal attribution certainly has predictive reliability: "it carries direct information about likely future events (the expected outcome) and its context" (Csibra–Gergely 2013). There is also evidence for the occurrence of teleology attribution that does not involve the attribution of mental states for the understanding and computation of actions by organisms (Csibra–Gergely 2013).

So, in terms of predictive power our teleological module seems to perform well enough. In this respect it is unlike crystal spheres and phlogiston, which means that the best reason for elimination is out. However, I don't want to argue here that teleological language should be interpreted as literally true in all of its applications or that it is the best tool for the description of living systems. What it seems to do is tracking the presence of self-determining systems and it can also be used as a tool in analysing the ways in which the parts of such systems achieve the supposed end of the whole. The ability of tracking the presence of self-determining systems and predicting their ends does not imply that teleolog-

ical language describes reality perfectly. Its applications can be faulty, and telos attributions can become unjustifiably loaded with e.g., anthropomorphic suppositions. However, any application can become subject of criticism and modified to match the features of a system more closely. The claim here is simply that, the function of teleological discourse is to track the presence of self-determining systems. Teleological discourse is not transparent in this respect, only science could tell us what are the underlying structural features that constitute such systems. This scientific language is more precise and should be considered more fundamental than teleological discourse. But the reduction suggested here still makes teleological discourse a respectable tool in describing reality, similarly to the use of the language of hardness or temperature.

But can't we still say that teleological discourse leads to a lot of confusion and misrepresentation? Let us start in the context of science and science education. According to many biologists it might create confusion by suggesting that evolution is a guided process, or by suggesting that organisms are designed by a creator, or by suggesting so-called Lamarckian mechanisms for evolution. To answer that worry let me point out this: teleological discourse, and our teleology seeking cognitive faculties are way older than our systematic accounts of the history of life, especially the theory of evolution. If we consider this faculty as a product of evolution, it must have had its uses in a context where humans didn't even consider the past more than a few generations before their own time. So, if teleological intuitions were selected for by evolution they were not selected for the capacity to grasp e.g. the mechanism of evolution, but probably for the capacity to track the behaviour of self-determining systems in the immediate environment. Therefore, it is not surprizing that they work fairly well for predicting the behaviour of other self-determining systems, but lead to misrepresentation when applied in a new context. So, there is a good reason to be cautious, but only for a restriction on the scope of application, not for eliminating the teleological module.

One should not forget that the same research quoted on the usefulness of teleological intuitions also shows, alongside other investigations into the subject, that humans have a strong tendency to overattribute teleology in other contexts as well (goals and functions alike). Teleology seems to work as a cognitive default heuristic. On the one hand, this means that infants and adults alike tend to attribute goals even to seemingly inanimate objects like rectangles, puppets, robots (Csibra 2008), basically if they behave in a sufficiently varied manner. On the other hand, especially young children, but also adults tend to choose teleological explanations even for purely physical occurences when they are asked to choose between purely physical-causal and teleological options (Kelemen 1999, Kelemen et al. 2012). Kelemen shows that children and, to a somewhat lesser extent, adults have a tendency to suppose even in the case of natural objects such as clouds, trees or mountains, that they serve a purpose. She named this

tendency 'promiscuous teleology', and contrasts it with the ability to use teleological reasoning in appropriate domains, what she calls 'selective teleology'. Humans are obsessed with goals from an early age, they look for them everywhere most probably because this facilitates social learning about instrumental action and problem solving (Csibra–Gergely 2013).

Is this a problem for someone who aims to show that teleology is a respectable property of natural systems? Not necessarily. In the literature on cognitive processing and decision-making there is differentiation between heuristic and systematic processing (Kahneman (2011) calls them system 1 and system 2). Heuristic processing is fast, and automatic whereas systematic processing is slow and effortful. The distinction is a result of idealization, and it is accepted that in many cases processing takes place in an in between manner, partly heuristic, partly systematic. Heuristics are useful simplifications that are helpful in the right kind of context, but systematically mislead in other contexts[2]. However, choosing systematic processing and gathering more information, we can correct heuristic biases. Kelemen et al. (2012) showed that time pressure, which increases the tendency to use simple heuristics, increases the occurrence of promiscuous teleology in test subjects. But it was also demonstrated that education reduces the occurrence of promiscuous teleology and the only factor that seemed to count was the level of education. It did not matter whether the subject had a PhD in literature or physics. Most probably the result is a consequence of more systematic processing and information seeking. Therefore, it is plausible to think, even if we are usually too obsessed with goals and overly reliant on one clue for identifying goal-directed, teleological behaviour (Csibra 2008), we have the means to correct ourselves and 'selective teleology' can be at least approximated.

Approaching the same issue from a different angle, even if there is a systematic bias in our teleological cognitive module towards false positives, that doesn't mean that the entity the module is searching for is a non-existent. So, the identification of a systematic bias is not a good enough reason for the elimination of teleological discourse. Still, the best explanation for the existence of the module is that it helps the organism to find or to avoid the entities modelled, described by it reliably, if not always correctly. Also, the fact that a cognitive module produces many false positives, without producing false negatives, is not an evolutionary disadvantage as long as the cost of false positives is low. Obviously, it is much less problematic to recognize a boulder as bear than the other way around. This is a well-recognized pattern in evolutionary psychology with respect to many cognitive categories. To save energy organisms manage errors

[2] For example, the ease of remembering is a good heuristic regarding the distribution of different weeds in my own garden, supposing that I know weeds and I visit my garden daily. The same heuristic is bad guide with respect to the frequency of suicides in my country.

only when those are costly enough, that is why this principle is called error management theory (see Buss 2016). To sum up, the ability to identify organisms, self-determining systems is a highly adaptive trait and the inability to do so is mostly a maladaptive one, regardless of the numerous low-cost false positives that are generated in the process (for a more detailed discussion of this topic see: Kertész–Kodaj 2023).

Before closing this paper, it might worth mentioning one further advantage of accepting the perspective argued for here. Namely, that teleological intuitions and language are mostly fair descriptions of reality and teleology (TE) as a property term can be reduced to the systemic property I called Organizational Closure (OC). As Kelemen, Csibra and others also recognized people tend to attribute teleology, over and above organisms, to things like social organizations, in some cases to groups. From an organizational point of view, I suspect that not all of these are cases of misattribution. The notion of organizational closure probably can be reasonably extended to include systems that are not organisms in the ordinary sense. Maybe the most obvious intuitive example would be the case of superorganisms, the large, well-organized colonies of social insects. But let's just take a look at artefact functions here which is usually taken to be unproblematic. But there is extensive tool use in the animal kingdom, the exquisite palaces of termites are a prime example. How could we handle them conceptually? According to received wisdom artefact function depends on the intentions of the maker. Do termites have intentions? Quite probably not.

Let's approach the case differently. The self-regulatory capacities of many organisms involve agential capacities, movement and in many cases the more or less complex manipulation of the environment. Following Moreno (2018), agency can be defined as changing one's environment in such a way that the change is beneficial or even indispensable from the point of view of the self-maintenance cycle of the agent. To use another example, a beaver cannot fulfil many of its basic needs without building its dam, it sounds plausible to say that the dam is a constraint constantly renewed by cycle that maintains the beaver and seems to serve as constraint that is important for cycle. If the dam can become part of the self-determination cycle of the beaver than in that cycle the dam serves an intrinsic function. Whether it does become part of it is a harder question to answer, the only thing I want to say here that this is not implausible. And this thought might open an interesting perspective on artefact functions. As in the case of termites we probably want to say that beavers don't have intentions. And think about this: you find and abandoned beaver dam or termite hill. Does it have a function? Well, according to the organizational perspective it might have had one, but at present it doesn't as it is not part of a self-determination cycle. These examples are mentioned to raise awareness, that the organizational account of teleology might be able to surprize us with unexpected solutions to old problems or even to unexpected problems connected to old problems.

In conclusion, we have good reasons to take teleological intuitions and teleological language seriously, giving teleological descriptions a realist reading. And the best way of taking them seriously is to suppose that these descriptions track the presence of self-determining systems, their behaviour and functioning.

REFERENCES

Ariew, A 2007. Teleology. In David L. Hull – Michael Ruse (Eds.) *The Cambridge Companion to the Philosophy of Biology*. Cambridge University Press.

Bedau, Mark 1992. Goal-Directed Systems and the Good. *The Monist*. 75/1. 34–51. https://doi.org/10.5840/monist19927516

Bertalanffy, Ludwig von 1951. General System Theory – A New Approach to Unity of Science. *Human Biology*. 23. 303–361.

Bertalanffy, Ludwig von 1968. *General Systems Theory*. George Braziller, New York

Bich, Leonardo – Bechtel, William 2021. Mechanism, Autonomy and Biological Explanation. *Biology and Philosophy*. 36/6. 1–27. https://doi.org/10.1007/s10539-021-09829-8

Breitenbach, Angela 2009. Teleology in Biology: A Kantian Perspective. *Kant Yearbook 2009*. 31–56.

Braithwaite, Richard B. 1953. *Scientific Explanation*. Cambridge, Cambridge University Press.

Buss, David M. 2016. *Evolutionary Psychology* (5th ed.). London, Routledge.

Csányi, Vilmos 1982. *A General Theory of Evolution*. Budapest, Akadémiai Kiadó

Csibra, Gergely – Gergely, György 2007. Obsessed with Goals: Functions and Mechanisms of Teleological Interpretation of Actions in Humans. *Acta Psychologica*. 124. 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

Csibra, Gergely – György Gergely 2013. Teleological Understanding of Actions. In Mahzarin R. Banaji – Susan A. Gelman (Eds.) *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*. Oxford, Oxford University Press. 38–43.

Churchland, Paul M. 1981. Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*. 78. 67–90. https://doi.org/10.2307/2025900

Crane, Tim 2001. *Elements of Mind*. Oxford, Oxford University Press

Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York, Oxford University Press

Dawkins, Richard 1976. *The Selfish Gene*. Oxford, Oxford University Press.

Dowe, Phil 2000. *Physical Causation*. New York, Cambridge University Press.

De Smedt, Johan – Helen De Cruz 2020. Is Intuitive Teleological Reasoning Promiscuous? In William Gibson – Dan O'brien – Marius Turda (Eds.) *Teleology and Modernity*. Abingdon and New York, Routledge. 185–202.

Fazekas, Péter 2009. Reconsidering the Role of Bridge Laws in Inter-Theoretical Reductions. *Erkenntnis*. 71/3. 303–322. https://doi.org/10.1007/s10670-009-9181-y

Garson, Justin 2019. *What Biological Functions Are and Why They Matter*. New York, Cambridge University Press.

Gelber, Jessica 2021. Teleological Perspectives in Aristotle's Biology. In Sophia M. Connell (eds.) *The Cambridge Companion to Aristotle's Biology*. Cambridge, Cambridge University Press. 97–113.

Ghiselin, Michael T. 1994. Darwin's Language May Seem Teleological, But His Thinking Is Another Matter. *Biology and Philosophy*. 9/4. 489–492. https://doi.org/10.1007/BF00850377

Godfrey-Smith, Peter 1994. A Modern History Theory of Functions. *Nous*. 28. 344–362. https://doi.org/10.2307/2216063

Kahneman, Daniel 2011. *Thinking, Fast and Slow*. London, Macmillan.
Kertész, Gergely – Kodaj, Dániel 2023. In Defense of Teleological Intuitions. *Philosophical Studies*. 180. 1421–1437. https://doi.org/10.1007/s11098-023-01937-3
Kim, Jaegwon 1992. Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*. 52/1. 1–26. https://doi.org/10.2307/2107741
Fazekas, Péter – Kertész Gergely 2011. Causation at Different Levels: Tracking the Commitments of Mechanistic Explanations. *Biology & Philosophy*. 26/3. 365–383. https://doi.org/10.1007/s10539-011-9247-5
Fazekas, Péter – Kertész Gergely 2019. Are Higher Mechanistic Levels Causally Autonomous? *Philosophy of Science*. 86. 847–857. https://doi.org/10.1086/705450
Hamilton, William D. 1964. The Genetical Evolution of Social Behaviour. *Journal of Theoretical Biology*. 7/1. 1–16. https://doi.org/10.1016/0022-5193(64)90038-4
Hartmann, Nicolai 1951. *Teleologisches Denken*. Berlin, Walter de Gruyter.
Kampis, György 1991. *Self-Modifying Systems in Biology and Cognitive Science*. Oxford, Pergamon Press.
Kelemen, Deborah 1999. Why Are Rocks Pointy? Children's Preference for Teleological Explanations of the Natural World. *Developmental Psychology*. 35. 1440–1452. https://doi.org/10.1037/0012-1649.35.6.1440
Kelemen, Deborah – Joshua Rottman – Rebecca Seston 2012. Professional Physical Scientists Display Tenacious Teleological Tendencies. *Journal of Experimental Psychology*: *General*. 142/4. 1074–83.
Lennox, James G. 1993. Darwin Was a Teleologist. *Biology & Philosophy*. 8/4. 409–421. https://doi.org/10.1007/BF00857687
Mayr, Ernst 1974. Teleological and Teleonomic: A New Analysis. In Robert S. Cohen – Marx W. Wartofsky (Eds.) *Methodological and Historical Essays in the Natural and Social Sciences (Boston Studies in the Philosophy of Science 14)*. Dordrecht, Springer Netherlands. 91–117.
Madrell, Simon H. P. 1998. Why Are There No Insects in the Open Sea? *The Journal of Experimental Biology*. 201. 2461–64. https://doi.org/10.1242/jeb.201.17.2461
Maturana, Humberto – Francisco Varela 1980. *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht, D. Reidel Publishing Company.
Maturana, Humberto – Francisco Varela 1992. *Tree of Knowledge: The Biological Roots of Human Understanding*. Boulder/CA, Shambhala Publications.
Mackie, John L. 1974. *The Cement of the Universe*. Oxford, Clarendon Press.
McLaughlin, Peter 2001. *What Functions Explain: Functional Explanation and Self-Reproducing Systems*. Cambridge, Cambridge University Press.
Millikan, Garrett R. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge/MA, MIT Press.
Mossio, Matteo – Giuseppe Longo – John Stewart 2009. A Computable Expression of Closure to Efficient Causation. *Journal of Theoretical Biology*. 257/3. 489–498. https://doi.org/10.1016/j.jtbi.2008.12.012
Mossio, Matteo – Leonardo Bich 2017. What Makes Biological Organisation Teleological? *Synthese*. 194/4. 1089–1114. https://doi.org/10.1007/s11229-014-0594-z
Moreno, Alvaro – Matteo Mossio 2015. *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Berlin, Springer. https://doi.org/10.1007/978-94-017-9837-2
Nagel, Ernest 1961. *The Structure of Science: Problems in the Logic of Explanation*. New York, Harcourt, Brace & World.
Nagel, Ernest 1970. Issues in the Logic of Reductive Explanations. In Howard E. Kiefer – Milton K. Munitz (Ed.) *Mind, Science, and History*. Albany/NY, SUNY Press. 117–137.
Rose, David – Jonathan Schaffer 2017. Folk Mereology is Teleological. *Noûs*. 51/2. 238–70. https://doi.org/10.1111/nous.12123

Savitt, Steven 1974. Rorty's Disappearance Theory. *Philosophical Studies*. 28. 433–36.
Schwartz, Peter H. 1999. Proper Function and Recent Selection. *Philosophy of Science*. 66. S210–S222. https://doi.org/10.1086/392726
Sklar, Lawrence 2015. Philosophy of Statistical Mechanics. *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition). https://plato.stanford.edu/archives/fall2015/entries/statphys-statmech/
Sommerhoff, Gerd 1969. The Abstract Characteristics of Living Systems. In Frederick E. Emery (Ed.) *Systems Thinking*. Middlesex, Penguin. 147–202.
van Inwagen, Peter 1990. *Material Beings*. Cornell University Press.
Walsh, Denis M. 2015. *Organisms, Agency, and Evolution*. Cambridge, Cambridge University Press.
Woodward, James 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford, Oxford University Press.

Erik Åkerlund

Models of Finality: Aristotle, Buridan, and Averroes

The treatment of the notion of finality has been a task ripe with pitfalls for historians of philosophy. Much criticism of the notions of final cause and final causation of earlier thinkers has been a criticism of a more contemporary conception of finality rather than a criticism of the notion as found in these earlier thinkers themselves. More specifically, much criticism of Aristotle's conception of the final cause is a critique of what later philosophers have read into the conception of finality rather than what Aristotle himself wrote about it and how he understood it.

Against this background, some kind of "heuristic scheme" is necessary in order to structure the different conceptions that different thinkers have had surrounding finality. In this article, one such scheme is worked out. The basic difference between the conceptions of finality presented in the following sections is that between an *intentionalist* and a *non-intentionalist* conception. According to an intentionalist conception of finality, a (rational) agent is necessary for there to be an end or an operation for an end. In a non-intentionalist conception, no such intentional agency is needed.

Connected to this concept, a distinction can also be made regarding different basic conceptions of the nature of reality, here called "metaphysics" for short,[1] namely between *Dynamic* and *Boolean* conceptions. A Dynamic conception of metaphysics is one in which change is understood on an act-potency scheme and in which things have potentialities that become actualized. In Boolean metaphysics, on the other hand, the radical division between that which is or exists on the one hand and that which does not exist on the other is emphasized; hence, change is rather understood as the replacement of one thing (a substance or an accident) with another.

[1] Thus, the "metaphysics" of a thinker can be found in the metaphysics as well as the natural philosophy and theology (and possibly also other subjects) of a thinker, depending on how the subjects are divided. Special attention will be given, though, to the understanding of reality as it pertains or relates to nature. When the word "metaphysics" is used in some other sense (e.g., as opposed to natural philosophy or theology), this will be noted.

It is argued here that these divisions shed light on the difference with regard to the problems that different thinkers face in relation to finality and that their proposed solutions will therefore also differ. To concretize the different treatments of finality, three different thinkers who have presented different combinations of the above views are first introduced. These are the following:

- Aristotle (non-intentionalist understanding of finality combined with Dynamic metaphysics);
- Buridan (intentionalist understanding of finality combined with Boolean metaphysics); and finally
- Averroes (intentionalist understanding of finality combined with Dynamic metaphysics).[2]

These three thinkers are not treated in chronological order. The reason for this, argued for in the presentation below, is that the two first combinations (Aristotle's and Buridan's) represent "cleaner" solutions and views regarding finality. The third view, on the other hand, seemingly introduces a tension or complication, as Dynamic metaphysics presents a kind of final explanation in itself (in terms of actualization of potentialities) that at least partly covers the same explanatory space as the intentionalist account of finality does. Hence, we are presented here with a question regarding how God's intentions and the actualization of potentialities relate to one another, in the end. Therefore, it is appropriate that Averroes is presented as the last thinker of the three.

Let us now consider the three thinkers in turn and examine how finality is embedded in their three respective world views.

I. CHANGE AND FINALITY IN ARISTOTLE

Let us start with Aristotle's *Physics* to understand what role finality[3] plays in his philosophy,[4] and especially in his philosophy of nature.[5] In his introduction to his treatment of causes *(aitiai)* in the *Physics*, Aristotle writes:

[2] The missing combination – a non-intentionalist understanding of finality combined with a Boolean metaphysics – would presumably amount to an eliminationist view of finality, i.e., that there is no genuine finality at all in the world. Ancient Epicureanism would presumably come close to this position.

[3] This term is here preferred over the term "teleology".

[4] On the importance of the notion of *telos* for Aristotle's philosophy and on its connection to the act-potency-scheme in his metaphysics, one recent commentator has concluded that "Aristotle can only really make sense of *ousia*, in relation to its basic intelligibility, through the concepts of *telos* (end) and *entelechy* (fulfilment)" (Brook 2015. 521).

[5] For a recent study on finality in Aristotle, with an overview over recent secondary literature, see Leunissen 2020.

> Knowledge is the object of our inquiry, and men do not think they know a thing till they have grasped the "why" of it (which is to grasp its primary cause). So clearly we too must do this as regards both coming to be and passing away and every kind of natural change, in order that, knowing their principles, we may try to refer to these principles each of our problems (Physics, book II, ch. 3; 194b17–23; Barnes 1984 vol. I. 332.)[6]

First of all, causes and causation have to do with an explanation of something, namely of change *(metabolē)*.[7] Thus, the fundamental question that causal accounts in natural philosophy are meant to answer is not of the kind "Why does this thing exist?" but rather "Why has this come into being?" or "Why did this thing change color?". Hence, causal explanation involves specifying the principles of change. Alternatively put, causation has to do with explaining change.

It is a salient feature of this entry into causation that it also leads to a certain "causal pluralism." Not in the sense that whichever explanation is a good explanation. But one good explanation does not preclude other good explanations. As Aristotle writes, "[a]s things are called causes in many ways, it follows that there are several causes of the same thing (not merely accidentally)" (*Physics*, book II, ch. 3; 195a4–5; Barnes 1984 vol. I. 333.).[8] The question "Why?" can be understood in many different ways, and so there are many different explanations for the same change.

It is also important to remember that we are now examining nature, as this is a book about nature *(physis)*; it is the "student of nature" who is to investigate the principles of change – change being significant for nature – and to make the proper distinctions with regard to the different meanings of "Why?".

> Now, the causes being four, it is the business of the student of nature to know about them all, and if he refers his problems back to all of them, he will assign the "why" in the way proper to his science – the matter, the form, the mover, that for the sake of which (*Physics*, book II, ch. 7; 198a22–24; Barnes 1984 vol. I. 338.).[9]

[6] "ἐπεὶ γὰρ τοῦ εἰδέναι χάριν ἡ πραγματεία, εἰδέναι δ᾽ οὐ πρότερον οἰόμετα ἕκαστον πρὶν ἂν λάβομεν τὸ διὰ τί περὶ ἕκαστον, τοῦτο δ᾽ ἐστὶ τὸ λαβεῖν τὴν πρώτον αἰτίαν, δῆλον ὅτι καὶ ἡμῖν υοῦτο ποιητέον καὶ περὶ γενέσεος καὶ φθοπᾶς καὶ πάσηες τῆς φυσικῆς μεταβολῆς, ὅπως εἰδότες αύῶν τὰς ἀρχὰς ανἀγειν είς αύτὰς πειρὠμεθα τῶν ζητουμένον ἕκαστον."

[7] In a recent study, Nathanael Stein (2011. 707n8) enumerates a list of scholars who have held that "Becauses," "explanations" or "explanatory factors" would be a better rendering of *aitiai* than "causes," mentioning i.a. Julia Annas, Jonathan Barnes, Richard Sorabji, and Michael Frede.

[8] "συμβαίνει δὲ πολλαχῶς λεγομένον τῶν αἰτίων καὶ πολλὰ τοῦ αὐτοῦ αἴτια εἶναι οὐ κατὰ συμβεβεκός."

[9] "ἐπεὶ δ᾽ αί αἰτίαι τέτταρες, περὶ παςῶν τοῦ φυσικοῦ εἰδέναι, καὶ εἰς πάσας ἀνάγων τὸ διὰ τί ἀποδώσει φυσικῶς – τὴν ὕλην, τὸ εἶδος, τὸ κινῆσαν, τὸ οὗ ἕνεκα."

However, although proper distinctions are to be made, this does not mean that the principles in the final analysis have to differ in the individual case. In the case of the paradigmatic example for Aristotle – that of the organism[10] – the last three of the above types of causes actually coincide.

> The last three often coincide; for the what and that for the sake of which are one, while the primary source of motion is the same in species as these (*Physics*, book II, ch. 7; 198a25–26; Barnes 1984 vol. I. 338.).[11]

Hence, it would seem that in organisms – which are "self-movers"[12] – the form (which determines what a thing is) is identical with the finality ("that for the sake of which") and to that which moves (or "the principle of motion").

At this stage as well, Aristotle balances on the limits of physics in an interesting way and delineates where that which transcends physics begins. This has to do with cases in which the example of a human being begetting another human being is changed in such a way that that which causes such a change does not itself move. What we are describing here, then, is the unmoved mover, or God:[13]

> [A]nd such as are not of this kind [i.e., those who do *not* move by being themselves moved] are no longer inside the province of natural science, for they cause motion not by possessing motion or a source of motion in themselves, but being themselves incapable of motion (*Physics*, book II, ch. 7; 198a28–29; Barnes 1984 vol. I. 338.).[14]

Here, we have a case in which the principle of change in or of nature is not itself part of nature and is therefore not the study object of the *Physics* but rather that of the *Metaphysics*, which we address below.

Hence – paradoxically, it would seem at first – the study of change in nature leads us both to principles that are themselves in nature (and therefore studied in the present context) and to principles that are not themselves part of nature (and therefore are *not* treated in the *Physics*). This is further stressed by Aristotle:

[10] See Shields 2014. 96.

[11] "ἔρχεται δὲ τὰ τρία εἰς [τὸ] ἓν πολλάκις· τὸ μὲν γὰρ τί ἐστι καὶ τὸ οὗ ἕνεκα ἕν ἐστι, τὸ δ' ὅθεν ἡ κίνησις πρῶτον τῷ εἴδει ταὐτὸ τούτοις·"

[12] Cf. Shields 2014. 324. "According to Aristotle, unlike artefacts, living systems engage in their activities *spontaneously*. He thinks that living beings are spontaneous in the sense that they have an internal source *(archê)* of change. While many things move, only some things are self-movers."

[13] This identification will be argued for below.

[14] "ὅσα δὲ μὴ, οὐκέτι φυσικῆς. οὐ γὰρ ἐν αὑτοῖς ἔχοντα κίνεσις οὐδ᾽ ἀρχὴν κινήσεως κινεῖ, ἀλλ᾽ ἀκίνητα ὄντα."

> Now the principles which cause motion in a natural way are two, of which one is not natural, as it has no principle of motion in itself. Of this kind is whatever causes movement, not being itself moved, such as that which is completely unchangeable, the primary reality, and the essence of a thing, i.e., the form; for this is the end for the sake of which. Hence since nature is for the sake of something, we must know this cause also. (Physics, book II, ch. 7; 198a36–198b5; Barnes 1984 vol. I. 338–9.)[15]

The principles that move without themselves being moved, then, are in Aristotle's short enumeration (i) the ultimate reality and (ii) the form *(morphē)* or essence *(ti esti)* of the thing. In the case of (i), moreover, the principle is described as "completely unchangeable," implicitly stating the ability to change of forms or essences (when things come into being or perish, presumably).

Aristotle's "causal pluralism" also comes to the fore in the above quote. To ask whether finality is internal *or* external to the thing misses the point.[16] As in the case with the four kinds of causes, it depends on which "Why?" question one is asking and on the way in which one understands it. Surely, the essence itself is a final cause of the thing. In change, the thing actualizes its potentialities, which are inherent in its form or essence. But change is also an actualization (without further qualifications), in which pure actuality is the unmoved mover. That which is fully actualized cannot actualize any "more" and cannot, therefore, change or be moved – hence it is an *unmoved* mover, which is "completely unchangeable."[17]

To make this point more fully, we must move outside physics – according to Aristotle himself – to the discipline that deals with this, the area later labeled metaphysics. The treatment in the *Metaphysics* is strikingly parallel in some parts to that in the *Physics*, but for that reason, the differences also stand out all the more clearly.

In the first chapter of book VI (or E), for example, Aristotle writes:

> We are seeking the principles and the causes of the things that are, and obviously of things *qua* being (*Metaphysics*, book VI, ch. 1; 1025b3–4; Barnes 1984 vol. II. 1619.).[18]

[15] "Διτταὶ δὲ αἱ ἀρχαὶ αἱ κινοῦσαι φυσικῶς, ὧν ἡ ἑτέρα οὐ φυσική. οὐ γὰρ ἔχει κινήσεως ἀρχὴν ἐν αὑτῇ. τοιοῦτον δ᾽ ἐστὶν ἔι τι κινεῖ μὴ κινούμενον, ὥπερ το τε παντελῶς ἀκίνετον καὶ τὸ πάντων πρῶτον καὶ τὸ τί ἐστι καὶ ἡ μορφή. τέλος γὰρ καὶ οὗ ἕνεκα. ὥστε ἐπεὶ ἡ φύσις ἕνεκά του, καὶ ταύτην εἰδέναι δεῖ."

[16] On this point, Monte Ransome Johnson (2005) is wide of the mark when he pits these against each other in his otherwise brilliant study on Aristotle's teleology (Johnson 2005. 284–6). It would seem that Johnson works from the presumption that "external" finality would have to involve some kind of "intentionalist teleology," which is not the case, as is seen precisely in Aristotle. For the term "intentionalist teleology," and for an argument for the view that Aristotle places himself between an eliminativist (Democritus et al.) and an intentionalist (Anaxagoras) stance on teleology, see Shields 2014, especially 86–87.

[17] See also *Metaphysics*, book V, ch. 1; $1012^{b}34$–$1013^{a}23$; Barnes vol. II. 1599–1600.

[18] "αἱ ἀρχαὶ καὶ τὰ αἴτια ζητεῖται τῶν ὄντων, δῆλον δὲ ὅτι ᾗ ὄντα."

Thus, that which Aristotle seeks to accomplish here is strikingly similar to his search for different kinds of causes in the *Physics*, except that here he goes beyond physics. For whatever Aristotle calls what he does here –despite his not using the word "metaphysics" – this is an investigation taking him beyond the physical realm:

> But if there is something which is eternal and immovable and separable, clearly the knowledge of it belongs to a theoretical science, not, however, to natural science (for natural science deals with movable things) nor to mathematics, but to a science prior to both (*Metaphysics*, book VI, ch. 1; 1026a10–13; Barnes 1984 vol. II. 1619.).[19]

Hence, we have come here to the principle, or principles, of the highest kind. These principles do not suspend the natural principles but are rather the principles on which the natural principles, are based in turn, at least partly (as music is partly "based" on mathematical principles).[20]

Here, we also find a textual basis for claiming that Aristotle understands himself as considering God or the divine here. He writes:

> There must, then, be three theoretical philosophies, mathematics, natural science, and theology *(theologikē)*, since it is obvious that if the divine is present anywhere, it is present in things of this sort. And the highest science must deal with the highest genus, so that the theoretical sciences are superior to the other sciences, and this to the other theoretical sciences. (*Metaphysics*, book VI, ch. 1; 1026a18–23; Barnes 1984 vol. II. 1619.)[21]

In addition, to clarify his position, Aristotle provides a counterfactual argument that if nature were everything that existed, natural science would be the highest science (as is indeed held by many, if not most, thinkers today[22]):

[19] "εἰ δέ τί ἐστιν ἀΐδιον καὶ ἀκίνητον καὶ χωριστόν, φανερὸν ὅτι θεωρητικῆς τὸ γνῶναι, οὐ μέντοι φυσικῆς γε (περὶ κινητῶν γάρ τινων ἡ φυσική) οὐδὲ μαθηματικῆς, ἀλλὰ προτέρας ἀμφοῖν."

[20] Here, though, Aristotle expressly writes that we search for principles of *things*, rather than of change, which is logical given the content of discourse, as physics deals with mutable things, whereas metaphysics also (or even exclusively?) deals with immutable things.

[21] "ὥστε τρεῖς ἂν εἶεν φιλοσοφίαι θεωρητικαί, μαθηματική, φυσική, θεολογική (οὐ γὰρ ἄδηλον ὅτι εἴ που τὸ θεῖον ὑπάρχει, ἐν τῇ τοιαύτῃ φύσει ὑπάρχει), καὶ τὴν τιμιωτάτην δεῖ περὶ τὸ τιμιώτατον γένος εἶναι. αἱ μὲν οὖν θεωρητικαὶ τῶν ἄλλων ἐπιστημῶν αἱρετώταται, αὕτη δὲ τῶν θεωρητικῶν."

[22] See, e.g., the article "Naturalism" in *Stanford Encyclopedia of Philosophy*, where it is stated that the term "naturalism" as used today is used to designate the view that "reality is exhausted by nature, containing nothing 'supernatural', and that the scientific method should be used to investigate all areas of reality, including the 'human spirit'." It goes on to state that "[s]o understood, 'naturalism' is not a particularly informative term as applied to contemporary philosophers. The great majority of contemporary philosophers would happily accept naturalism as just characterized – that is, they would both reject "supernatural" entities, and

> We answer that if there is no substance other than those which are formed by nature, natural science will be the first science; but if there is an immovable substance, the science of this must be prior and must be first philosophy, and universal in this way, because it is first (*Metaphysics*, book VI, ch. 1; 1026a27–31; Barnes 1984 vol. II. 1620.).[23]

This is important because, as was seen in the *Physics*, this highest principle also comes into use when addressing natural change, in order to explain it. And the highest principle has this explanatory role also with respect to finality in nature.

Let us finally consider how human beings enter into this scheme. In the course of asking about the good of or for human beings, Aristotle places human beings exactly in the view of nature and other natural things:

> Life seems to be common even to plants, but we are seeking what is peculiar to man. Let us preclude, therefore, the life of nutrition and growth. Next there would be a life of perception, but *it* also seems to be common even to the horse, the ox, and every animal. There remains, then, an active life of the element that has a rational principle. (*Nicomachean ethics*, book I, ch. 7; 1097b34–1098a4; Barnes 1984 vol. II. 1735.)[24]

Of particular interest here is how Aristotle searches for the finality of human beings in line with how other things in nature fulfill their ends, namely by realizing that which is proper to them. That which is most proper to human beings is their rationality, and living in accordance with this, realizing this, is to live a good life (or the best life).

However, there are of course dissimilarities as well to how other entities in nature realize their capacities. This end is – in one sense – already realized in human beings. Placing the *Nicomachean ethics* next to the *De anima*, one can find this remarkable feature of human life here and now:

> Thought in this sense of it is separable, impassable, unmixed, since it is in its essential nature activity (for always the active is superior to the passive factor, the originating force to the matter).
>
> Actual knowledge is identical with its object; in the individual, potential knowledge is in time prior to actual knowledge, but absolutely it is not prior even in time. It does not sometimes think and sometimes not think. When separated it is alone just what it

---

allow that science is a possible route (if not necessarily the only one) to important truths about the 'human spirit'" (Papineau 2016).

[23] "εἰ μὲν οὖν μὴ ἔστι τις ἑτέρα οὐσία παρὰ τὰς φύσει συνεστηκυίας, ἡ φυσικὴ ἂν εἴη πρώτη ἐπιστήμη: εἰ δ᾽ ἔστι τις οὐσία ἀκίνητος, αὕτη προτέρα καὶ φιλοσοφία πρώτη, καὶ καθόλου οὕτως ὅτι πρώτη"

[24] "τὸ μὲν γὰρ ζῆν κοινὸν εἶναι φαίνεται καὶ τοῖς φυτοῖς, ζητεῖται δὲ τὸ ἴδιον. ἀφοριστέον ἄρα τήν τε θρεπτικὴν καὶ τὴν αὐξητικὴν ζωήν. ἑπομένη δὲ αἰσθητική τις ἂν εἴη, φαίνεται δὲ καὶ αὐτὴ κοινὴ καὶ ἵππῳ καὶ βοῒ καὶ παντὶ ζῴῳ. λείπεται δὴ πρακτική τις τοῦ λόγον ἔχοντος."

> is, and this alone is immortal and eternal (we do not remember because, while this is impassible, passive thought is perishable); and without this nothing thinks. (*De anima*, book III, ch. 5; 430a17–25; Barnes 1984 vol. I. 684.)[25]

Many of the properties accorded to the active intellect as presented in the above quotations are identical to those of the unmoved mover or highest principle. We need not establish the exact relationship between these in this context; suffice it to say that because rational thinking is most proper to human beings, being their end and constituting (at least partly) that which is truly good for them, they also stand in a special relation to the overarching good, that highest principle which is – in one sense – the ultimate end of everything (not precluding, as has been stated, that all things also have their own internal proper end as well).[26]

The main point of this sketch is to present the broad lines on which Aristotle has presented a concept of finality that does not, at its core, draw on an intentionalist understanding of this phenomenon.[27] Rather, intentionality in general, and rational intentionality in particular, are placed within this more overarching idea of finality, which involves the explanation of change and in which change is fundamentally understood as actualization of potentialities.

II. BURIDAN, METAPHYSICS AND FINALITY

Buridan famously never left the arts department for higher theological studies. However, some parts of his philosophy were shaped by close contact with theological themes. One such area of his philosophy is his understanding of the status of accidental forms and, thereby, his understanding of change and motion.

As Paul Bakker has convincingly argued, Buridan's discussion of the status of accidental forms is very much informed by theological concerns.[28] More specif-

[25] "καὶ οὗτος ὁ νοῦς χωριστὸς καὶ ἀπαθὴς καὶ ἀμιγής, τῇ οὐσίᾳ ὢν ἐνέργεια. ἀεὶ γὰρ τιμιώτερον τὸ ποιοῦν τοῦ πάσχοντος καὶ ἡ ἀρχὴ τῆς ὕλης. τὸ δ' αὐτό ἐστιν ἡ κατ' ἐνέργειαν ἐπιστήμη τῷ πράγματι. ἡ δὲ κατὰ δύναμιν χρόνῳ προτέρα ἐν τῷ ἑνί, ὅλως δὲ οὐδὲ χρόνῳ, ἀλλ' οὐχ ὁτὲ μὲν νοεῖ ὁτὲ δ' οὐ νοεῖ. χωρισθεὶς δ' ἐστὶ μόνον τοῦθ' ὅπερ ἐστί, καὶ τοῦτο μόνον ἀθάνατον καὶ ἀΐδιον (οὐ μνημονεύομεν δέ, ὅτι τοῦτο μὲν ἀπαθές, ὁ δὲ παθητικὸς νοῦς φθαρτός). καὶ ἄνευ τούτου οὐθὲν νοεῖ."

[26] For an account of human mental activity as related to motion and change in Aristotle, see Shields 2007. There, Shields also points to some problems Aristotle ends up with in his account of this, in relation to his general account of change and action (see especially 159–160).

[27] Indeed, the highest principle is itself though (*nous*), according to Aristotle. However, this is only thought on thought itself (i.e., it thinks itself), and does not think about something else. Hence, this thought is not *per se* involved in some "directing" of the natural world, as it is, as an intentionalist understanding of ends in nature would have it. "Therefore it must be itself that thought thinks (since it is the most excellent of things), and its thinking is a thinking on thinking." (*Metaphyiscs*, book XII, ch. 9; 1074^{b}33–35; Barnes 1984 vol. II. 1698. 'αὐτὸν ἄρα νοεῖ, εἴπερ ἐστὶ τὸ κράτιστον, καὶ ἔστιν ἡ νόησις νοήσεως νόησις.')

[28] Bakker 2001, especially 252–3.

ically, it is the doctrine of the Eucharist, and the preservation of the accidental forms despite the change of the substantial form in it, that in a large part drives his discussion in this area.[29] The resulting view is one that could be called a "substantialization of accidental forms", as accidental forms are able to exist in this view without inhering in a subject.[30]

> In contrast to the Aristotelian conception of the ontological status of accidental being, Buridan offers a different theory by taking into account the point of view of the faith. Referring explicitly to the subsistence of the Eucharistic accidents *sine subiecto*, he takes his point of departure in an affirmation of Gods [sic] power to separate accidents from their substances. From this, he deduces that whiteness, in order to exist on its own *(per se)*, must be a real being, and hence that it possesses the status of a being not only while existing separately from a substance, but also while inhering in a substance. (Bakker 2001. 252–253.)[31]

This understanding of the status of accidental forms is combined in Buridan's view with an emphasis on the distinction between that which exists on the one hand and that which does not exist on the other. This can be seen, for example, in his arguments for the actuality of prime matter:

> The second conclusion is that [prime matter] is a being in act, not only in potency, because to be in potency only is not to be, but to be possible; but prime matter not only can be, but is, as was said (Buridan, *In Physicorum*, bk. I, q. 20; 202).[32]

Indeed, for matter to be able to account for anything, it needs to be real; it has to exist. As Buridan writes on causes in general:

[29] For this question as it regards Buridan, see also Sylla's (2001) contribution in the same anthology. The impact on philosophy from questions concerning the Eucharist, especially on the question of the status of accidental forms, is an important factor in the philosophy of the Late Middle Ages at large. Buridan was quite typical, then, in this respect. See Pasnau 2011, especially chs. 10 and 20.

[30] There is, of course, no necessity in this conclusion. In Marsilius of Inghen, for one, accidental forms do not acquire this status. This comes at the price, though, of more clearly separating the fields of natural philosophy, on the one hand, from the field of theology, on the other. Marsilius of Inghen thus upholds a distinction, closer to Aristotle, between substances (*quid*), on the one hand, and accidents as modifications of substances (*quale)*, on the other, in his natural philosophy. (Cf. Bakker 2001. 257–262).

[31] The quote goes on to note that Buridan has a univocal understanding of the term "being."

[32] "Secunda conclusio est quod ipsa est ens in actu, non solum in potentia, quia esse solum in potentia non est esse, sed posse esse; materia autem prima non solum potest esse, sed est, ut dictum est."

> That which is nothing is the cause of nothing (Buridan, *In Physicorum*, bk. II, q. 7; 294).[33]

Together with the understanding of the status of accidental forms, one can here see how accidental change is not so much the realization of a potentiality as it is the successive replacement of one accidental form with another. In line with accidents becoming more like substances, accidental change also becomes more like substantial change (or generation and corruption).

Calvin Normore has accounted for the difference between Buridan and Aristotle's account of change in the following way:

> Aristotle allows three kinds of change: generation/corruption, alteration, and motion. We can conceive of these in two fundamentally different ways. The first way (Aristotle's way!) is to think of them as different kinds of processes which a single thing, Socrates say, may in some sense suffer: Socrates was born, can move, can change size, can die. A second way is to conceive the different types of change as involving the creation and destruction of different kinds of things – in generation and destruction substances are created and destroyed, in movement, motions, in alterations, qualities, quantities, etc. This second picture does in a sense unify change by bringing them under the description "creation or destruction of something". Professor Adams has argued that Ockham understands the project of accounting for change in accord with the first picture. I want to argue that Buridan, on the other hand, is guided by the second and that this leads him to multiply entities but reduce modes. (Normore 1985. 195–196.)

In this unified understanding of change, connected to a univocal conception of being, all change is characterized by the destruction of one being and the introduction of another.[34] Once again, this view also precludes Aristotle's understanding of change as the realization of potentialities, as the potential simply does not exist and therefore cannot have any role to play in a causal explanation.

This is also connected to Buridan's understanding of modalities, and specifically on unrealized possibilities. Buridan underscores not only that unrealized possibilities have no existence whatsoever but also – and here we are entering the questions of final causation – that talk of unrealized possibilities only makes sense in relation to agents with free will. As Simo Knuuttila has written:

> As for the unrealized possible beings *(possibilia)*, Buridan states that they have no kind of existence and are not founded on anything (Knuuttila 2001. 71).

[33] "Quod nihil est nullius est causa."

[34] For Buridan's understanding of change, see also his *Super octo libros De generatione et corruption*, bk. I, qq. 6–9 (Buridan 2010. 77–104).

> In describing the behavior of created things, the notion of unrealized alternative possibilities is relevant only with respect to agents which have a free will (Knuuttila 2001. 72).

We have thus entered the realm of final causation. As Henrik Lagerlund has pointed out, final causation only occurs when there is a rational agent, according to Buridan.

> Obviously nothing in nature acts for the sake of the good other than humans. […] Ends are just intentions of rational agents (Lagerlund 2011. 600).

It remains, though, for Buridan to provide an account of this final causation and how it fits with the other causes.[35] Buridan does this by distinguishing what he calls "first intentions" from "second intentions."[36] When someone performs an act for an end, we can distinguish two senses of the word "end": (i) the end in the sense of the one *for the sake of* which (or whom) the action is performed *(finis gratia cuius)* on the one hand and (ii) the end in the sense of that *through* which something is achieved *(finis quo)* on the other. The end in the first sense is the primary sense of the word "end," and it is only here that we find true final causation.[37] The end in the secondary sense is rather the result of efficient and formal causation.[38]

> It is therefore to be conceded that an end said in first intention is truly a cause […] But it is also to be conceded that it is not fitting that an end said in the second intention is, properly speaking, a cause of its agents or the acts preceding it […]. (Buridan, *In Physicorum*, bk. II, q. 7; 298.)[39]

[35] For a recent, general account of Buridan on finality, see Pasnau 2020.

[36] In *In Physicorum*, bk. II, q. 7; see especially 296–298.

[37] Suárez would call this kind of end the *finis cui*, "the end for whom," reserving the term *finis cuius* for something coming close to Buridan's *finis quo* (confusingly, in this context). See, for example, *DM* XXIII.2, 2. "nam finis *cuius* dicitur cuius adipiscendi gratia homo movetur et operator, u test sanitas in curatione; finis *cui* dicitur ille cui alter finis procurator, ut test homo in intentione sanitas".

[38] Cf. Lagerlund 2011, especially 596–600. "It is ends in the second sense that Buridan dismisses since they come about through ends in the first sense, which means that they are effects and not causes." (Lagerlund 2011. 598)

[39] "Sic igitur concedendum est quod finis prima intentione dictus vere est causa […] Sed concedendum est etiam quod non oportet finem secundaria intentione dictum esse proprie loquendo causam suorum agentium vel actionum praecedentium ipsum […]."

> I therefore declare that the intention and will of the physician willing to heal Socrates does not depend on the coming about of Socrates' health. First, because this is nothing. Second, because it might be impossible for Socrates to be healed. (Buridan, *In Physicorum*, bk. II, q. 13; 345.)[40]

Thus, it is important to note that Buridan does not reject final causation but rather that he accepts final causation of a very specific kind, connected to rational agents and rational agency.

> The end in the first intention *(prima intentione)* is that which is first in the order of being, goodness, and perfection. It is that for which, or for the sake of which *(gratia cuius)*, something or someone acts. For example, it can be the man for whom the house is constructed. If we consider the whole universe, it is God who is in this sense the end of everything. (Biard 2001. 86–7)

Final causation, then, cannot really be used to explain what takes place within nature (outside human agency, one could add).[41] Here, other kinds of causes are in play.

> But as far as natural things are concerned, I believe that a swallow mating, nesting, and laying eggs does not cognize any more when it produces chicks than a tree does when it produces branches and flowers. Nor do the mating, nesting, and egg-laying activities of the swallow depend for their being and order on those chicks. Rather, the converse is true. And those chicks do not determine the swallow to act in this way, but the form and nature of the swallow, celestial bodies at certain times of the year, and supreme God in his infinite wisdom, together determine the swallow to mate, from which the production of eggs consequently follows. […] All of this comes about

[40] "Declaro igitur quod intentio et voluntas medici volentis sanare Socratem non dependet ex sanitate Socratis producenda. Primo, quia illa nihil est. Secundo, quia forte impossibile est Socratem sanari."

[41] There is a problem here, though, that Buridan does not seem to address. For if a cause has to exist in order to be a cause, how do we understand the situation in which a doctor is motivated to heal a patient that actually does not exist? In the above example, on Socrates, the patient does exist. But it could be the case that the doctor falsely assumes someone to exist, and is motivated in his or her actions to heal this person. The there is nothing that takes the place of the final cause. Hence, should we rather say that it is the *mental conception*, or something like that, that takes the place of the final cause, rather than the thing itself (e.g. Socrates)? It is questions and worries such as these that will shape the debate on final causation in the later Middle Ages. It should be noted, though, that in the most important case – that of God directing everything toward Himself as a first intention, this worry is not present, as God does exist (and if He didn't, He would not act, so the problem would not be present).

> by divine artifice, celestial bodies, and particular agents, both extrinsic and intrinsic [to the subject of the action], which are the substantial forms of these same natural agents. (Biard 2001. 88)[42]

Thus, even though God ultimately creates and upholds everything for a final end, it is formal and efficient causation that are the relevant causal categories when studying nature.

III. AVERROES ON ENDS, GOD'S AGENCY, AND THE ACT-POTENCY SCHEME

In Averroes, we find on the one hand an evaluation of God as an intentional agent, freely creating and upholding the world and endowing it with its structure and therefore also its ends, and on the other hand an analysis of at least the sub-lunar world in accordance with the Aristotelian four-fold scheme of causes, thereby also incorporating final causes understood on the act-potency scheme.

Exactly how the different parts of Averroes' philosophy and theology do or do not fit together is of course the subject of considerable debate and has been since his own lifetime.[43] However, it suffices to argue in this context that Aver-

[42] Translation of *In Physicorum*, bk. II, q. 7 (page 347 of the edition used here, based on slightly different text variant). "Sed de naturalibus ego credo, quod hirundo coiens, nidificans et ovificans nihil plus cognoscit pullos generandos quam arbor fronds et florens producens cognoscit fructum generandum. Nec hirundinis coitum, nidificatio et ovidificatio dependent in esse et ordine eorum ab illis pullis sed e contra. Nec illi pulli determinant hirundinem ad sic operandum, sed forma et natura hirundinis et corpora celestia determinatis temporibus et Deus supremus per suam sapientiam infinitam determinant hirundinem ad coitum, ex quo consequenter sequitur generatio ovorum […] Haec ergo omnia proveniunt ab arte divina et corporibus caelestibus et agentibus particularibus tam extrinsecis qum intrinsecis, quae sunt formae substantiales ipsorum naturalium."

[43] Barry Kogan, for example, in his *Averroes and the Metaphysics of Causation* (1985), suggests that there is an esoteric reading of Averroes, that can be extracted if more independent works such as the *Tahāfut al-tahāfut* are combined with the readings of Averroes' commentaries on Aristotle, especially the later, longer commentaries. (See especially page 24 and, for a summary of Averroes' "real" view in four points, page 232.) Oliver Leaman, on the other hand, in his book *Averroes and His Philosophy*, expressly rejects an esoteric reading (Leaman 1988. 127–128), and he argues that the traditional reception of "Averroism" in the Latin west represents a fairly accurate understanding of Averroes' own views (see especially 104 and 163–164). However, according to Leaman, the inherent tensions in Averroes' thinking were not as acute for him as it would become in a later Latin context, as philosophy and theology for Averroes was more about the organization of a good society, and about leading a good life, than about coming to an abstract, theoretical truth *per se* (144, 167–169, the latter with a comment on Pomponazzi's reception of Averroes). Furthermore, terms used in philosophy and theology, respectively, are used analogically (or equivocally *pros hen*), as they are used in different contexts and for different purposes, and so seeming inconsistencies between these two areas are only surface phenomena (183–184, 196). Leaman makes a strong case for his reading, but accepting it will also make the project of understanding Averroes' view (in singular) on

roes does in fact combine what is here called dynamic metaphysics with an intentionalist understanding of finality, although the details of this combination have been omitted in this commentary.[44]

To specify further, Averroes analyzes change in the sub-lunar world in accordance with the four-cause scheme. Hence, there are ends in nature that can be understood on the act-potency scheme. These ends, viewed only in themselves, are something inherent in the things.[45] If understood under the name "wisdom," as Averroes sometimes does, the things have this "wisdom" in themselves.

> For the philosophers believe that there are four causes: agent, matter, form, and end (Averroes, *Tahāfut al-tahāfut*, Discussion three; vol. I. 89–90; 150:15–151:8).[46]

> His [i.e., al Ghazālī's] assertion that not every cause is called an agent is true, but his argument that the inanimate is not called an agent is false, for the denial that the inanimate exhibits acts excludes only the rational and voluntary act, not act absolutely, for we find that certain inanimate things have powers to actualize things like themselves; e.g. fire, which changes anything warm and dry into another fire like itself, through converting it from what it has in potency into actuality. (Averroes, *Tahāfut al-tahāfut*, Discussion three, vol. I. 92; 154: 8–14.)

> [W]hen one observes this sublunary world, one finds that what is called "living" and "knowing" moves on its own account in well-defined movements towards well-defined ends and well-defined acts from which new well-defined acts arise (Averroes, *Tahāfut al-tahāfut*, Discussion three, vol. I. 112–113; 187:15–17.).

---

some *one* issue problematic. To these two readings can also be added a third, later one, from Ruth Glasner in her *Averroes' Physics* (2009). In this, she rather tries to show a development in Averroes' physics, where what she calls an "Aristotelian atomism" (first mentioned on page 2) is developed over time. This reading, if accepted, ought also to have repercussions on the understanding on Averroes' natural philosophy as a whole, including the status of final causes and final causation.

[44] The basis for the interpretation of Averroes' view here will be his *Tahāfut al-tahāfut* ("The Incoherence of the Incoherence"), written in his "middle period" (cf. Urvoy 1991. 36–38). This work, written in the late 1170s in response to al-Ghazālī's *Tahāfut al-Falāsifa* ("The Incoherence of the Philosophers"), is often taken as an expression as Averroes' own view. See, e.g., Kogan 1985. ix, Leaman 1988. 10, Urvoy 1991. 71.

[45] One can here also note how, for Averroes, potency or potentiality precedes possibility, where the possible is grounded in the actual (Leaman 1988. 29). Averroes develops this view in explicit opposition to Avicenna and al-Ghazālī, for whom the possible precedes the potential. Averroes develops his view in continuation with Aristotle and his "principle of plenitude", where everything that is possible will also at some point be realized. (For the principle of plenitude in Averroes, and the interpretative history of this with regards to Aristotle's philosophy, see Kukkonen 2000, especially 336n 23.)

[46] The last item in the reference refers to Bouyges' edition, in the *Bibliotheca Arabica Scholasticorum* series, vol. iii, Beyrouth, 1930.

One of the most interesting arguments for the view that there is indeed real causation taking place among things in the world is that without real causation in the world, knowledge would be impossible. For we come to know of things' natures through their operations – we do not have any "direct insight" into the nature of things. Hence, if the operation – or real causation – of things were to be denied, one would also have to deny the possibility of coming to know them.

> That a stone moves downward through a quality which has been created in it, and fire upwards, and that these qualities are opposed – this is a self-evident fact, and to contradict it is pure folly. But it is still more foolish to say that the eternal Will causes the movement in these things everlastingly – without any act He deliberately chose – and that this movement is not implanted in the nature of the thing, and that this is called constraint; for if this were true, things would have no nature, no real essence, no real definition at all. For it is self-evident that the natures and definitions of things only differ through the difference of their acts. (Averroes, Tahāfut al-tahāfut, Discussion 14, 289; 475:4–11.)

Averroes believes that causation and, thereby, the act-potency scheme or structure are actually "laid down" in the things themselves; they are not merely extrinsic to them, on account of God's agency. In line with this idea, Averroes also often underscores the indirect way in which God operates in the sub-lunar world. This agency in many ways takes place primarily through the heavens, which are themselves endowed with knowledge as well as will:

> As to the second hypothesis, that God moves the heavens without having created a potency in them through which they move, this also is a very reprehensible doctrine, far from man's understanding. It would mean that God touches and moves everything which is in the sublunary world, and that the causes and effects which are perceived are all without meaning, and that man might be man through another quality than the quality God has created in him and that the same would be true of all other things. But such a denial would amount to a denial of the intelligibles, for the intellect perceives things only through their causes. (Averroes, *Tahāfut al-tahāfut*, Discussion 14, 291; 479:1–7.)

> And this is one of the arguments through which it is established that the heavenly bodies are provided with intellect and desire; and this is clear also from various other arguments (Averroes, *Tahāfut al-tahāfut*, Discussion 14, 292; 480:16–18.).

Here, we come closer to the question of final causation specifically, for it seems that things in the sublunary world act for ends they possess in and of themselves. However, more proximately than from God, these ends are indicated by the heavens and the way in which these – the living heavenly bodies – move the

world. With a division that Buridan would later employ,[47] Averroes makes a distinction between the end of the heavens as a first intention – which is God – and the end of the heavens as a second intention – as they give ends to the sublunary world:

> This movement, however, does not occur according to the philosophers in first intention for the sake of this sublunary world; that is, the heavenly body is not in first intention created for the sake of this sublunary world. For indeed this movement is the special act for the sake of which heaven is created, and if this movement occurred in first intention for the sake of the sublunary world, the body of the heavens would be created only for the sake of this sublunary world, and it is impossible, according to the philosophers, that the superior should be created for the sake of the inferior. (Averroes, *Tahāfut al-tahāfut*, Discussion 15, 295; 484: 13–18.)

> This theologian [i.e., al Ghazālī] wants to indicate the cause of this from the point of view of the final cause, not of the efficient, and none of the philosophers doubts that there is here a final cause in second intention, which is necessary for everything in the sublunary world. And although this cause has not yet been ascertained in detail, nobody doubts that every movement, every progression or regression of the stars, has an influence on sublunary existence, so that, if these movements differed, the sublunary world would become disorganized. But many of these causes are either still completely unknown or become known after a long time and long experience, as it is said that Aristotle asserted in his book *On Astrological Theorems*. (Averroes, *Tahāfut al-tahāfut*, Discussion 15, 299; 491:13–492:5.)

Thus, the sublunary world – operating in accordance with an act-potency scheme – is given its ends and its general ordering more proximately from the heavens, which operate in accordance with reason and desire. However, it is of course ultimately from God that the ends, the structure, and the ordering come.[48]

> It also becomes clear from the fact that all the spheres have the daily circular movement, although besides this movement they have, as the philosophers had ascertained, their own special movements, that He who commands this movement must be the First Principle, i.e. God, and that He commands the other principles to order the other movements to the other spheres. Through this heaven and earth are ruled as a state is ruled by the commands of the supreme monarch, which, however, are transmitted to all classes of the population by the men he has appointed for this purpose in the different affairs of the state. As it says in the Koran: "And He inspires every Heaven with

[47] See above.

[48] And it is here, then, that one can find the break with Aristotle. God does not think about anything other than himself, according to Aristotle, whereas this is the case in Averroes.

> its bidding." This heavenly injunction and this obedience are the prototypes of the injunction and obedience imposed on man because he is a rational animal. (Averroes, *Tahāfut al-tahāfut*, Discussion three; vol. I. 111–112; 185:12–186:5.)

Above, we have seen how Averroes combines an act-potency scheme in his analysis of nature with a fundamentally intentionalist understanding of finality, or – more precisely – an analysis in which the end must ultimately be provided by a rational agent.[49] This concept would then be what is called here a *Dynamic* metaphysics with an *intentionalist* understanding of finality.[50]

It does not seem that Averroes problematizes this specific combination anywhere. However, whereas the combinations found in Aristotle and Buridan and presented above represent more "clean" solutions to how metaphysics and finality can be combined, it seems that with Averroes' combination, we have a situation in which two different accounts compete for the same "explanatory space." When accounting for a change in terms of the end, we can either explicate it along more traditionally Aristotelian lines as the actualization of a potentiality or we can refer it to the will of some rational agent (to a celestial agent or to God). Although a basic answer to the question of how these different accounts can be combined could be given along the lines of the *Liber de causis*, with its distinction between first order and second order causation, there is a tension in this account of the end that is not present in Aristotle's or Buridan's thinking. This tension, and the questions that it prompts, would later play a major role in the developments of the Latin and, more broadly, the "Aristotelian" philosophical traditions.

IV. CONCLUSIONS

Whereas Aristotle understands finality in a non-intentionalist way as the actualization of a potentiality, for Buridan, finality only emerges from the operation of a rational agent. In Averroes, the act-potency scheme used to explicate the workings of especially the sub-lunar world and its ends is combined with an intentionalist understanding of finality, in which the entire order of the world is ultimately dependent on the intentions and commands of God.

We can thus derive the following "four-field matrix":

[49] With all the *caveats* given above of how to exactly understand his combination of the philosophical and the theological perspectives.

[50] See also Cerami 2015, especially the conclusion on 672–675.

Table 1. Four-field matrix of different conceptions of reality and conceptions of finality.

		Conception of finality	
		Non-intentionalist	*Intentionalist*
Conception of reality	*Dynamic*	Aristotle	Averroes
	Boolean	—	Buridan

In the later Latin Middle Ages, a purely Aristotelian concept of finality was not truly accessible. Thus, what we have are understandings of the question that oscillate around "Buridanean" or "Averroist" expressions and solutions, in the sense of combining a basically intentionalist understanding with Dynamic metaphysics (in which ends can also be understood to be inherent in nature) or Boolean metaphysics (in which the ends tend to be understood as being extrinsic to things in the world).

Hopefully, the above heuristic scheme can serve to explain some of the different ways in which the notion of finality can be embedded into different world views and philosophies and also explain why different questions, problems, and challenges have been raised in relation to this notion for different thinkers in the history of philosophy.

REFERENCES

Averroes 1954. *Tahafut al-Tahafut*. Trans. Simon van den Bergh. London, Luzac.

Bakker, Paul J. J. M. 2001. Aristotelian Metaphysics and Eucharistic Theology: John Buridan and Marsilius of Inghen on the Ontological Status of Accidental Being. In Thijssen–Zupko 2001. 247–264. https://doi.org/10.1163/9789004452893_016

Barnes, Jonathan 1984. *The Complete Works of Aristotle*. Princeton/NJ, Princeton University Press.

Biard, Joël 2001. The Natural Order in John Buridan. In Thijssen–Zupko 2001. 77–95. https://doi.org/10.1163/9789004452893_008

Brook, Angus 2015. Substance and the Primary Sense of Being in Aristotle. *The Review of Metaphysics*. 68/3. 521–544.

Buridan, Jean 2015. *Quæstiones Super Octo Libros Physicorum Aristotelis (Secundum Ultimam Lecturam): Libri I–II*. Ed. Michiel Streijger – Paul J. J. M. Bakker. Leiden, Brill.

Buridan, Jean 2010. *Quaestiones Ssuper Octo Libros De Generatione et Corruptione Aristotelis: A Critical Edition with an Introduction*. Ed. Michiel Streijger – Paul J. J. M. Bakker – Hans J. Thijssen. Leiden, Brill.

Cerami, Cristina 2015. *Génération et Substance. Aristote et Averroès entre physique et métaphysique*. Boston, de Gruyter.

Glasner, Ruth 2009. *Averroes' Physics*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567737.001.0001

Johnson, Monte Ransome 2005. *Aristotle on Teleology*. Oxford, Oxford University Press.

Knuuttila, Simo 2001. Necessities in Buridan's Natural Philosophy. In Thijssen – Zupko 2001. 65–76. https://doi.org/10.1163/9789004452893_007

Kogan, Barry S. 1985. *Averroes and the Metaphysics of Causation*. Albany, State University of New York Press.

Kukkonen, Taneli 2000. Possible World in the *Tahāfut al-Tahāfut*: Averroes on Plenitude and Possibility. *Journal of the History of Philosophy*. 38/3. 329–347.

Lagerlund, Henrik 2011. The Unity of Efficient and Final Causality: The Mind/Body Problem Reconsidered. *British Journal for the History of Philosophy*. 19/4. 587–603. https://doi.org/10.1080/09608788.2011.583413

Leaman, Oliver 1988. *Averroes and His Philosophy*. Oxford, Clarendon Press. https://doi.org/10.4324/9781315027050

Leunisse, Mariska 2020. Teleology in Aristotle. In McDonough 2020. 39–63. https://doi.org/10.1093/oso/9780190845711.003.0003

McDonough, Jeffrey K. 2020. *Teleology: A History*. Oxford, Oxford University Press. https://doi.org/10.1093/oso/9780190845711.001.0001

Normore, Calvin 1985. Buridan's Ontology. In his *How Things Are. Studies in Predication and the History of Philosophy and Science*. Dordrecht, D. Reidel. 189–203. https://doi.org/10.1007/978-94-009-5199-0_8

Papineau, David 2016. Naturalism. In Edward N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/naturalism/

Pasnau, Robert 2011. *Metaphysical Themes: 1274–1671*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567911.001.0001

Pasnau, Robert 2020. Teleology in the Later Middle Ages. In McDonough 2020. 90–115. https://doi.org/10.1093/oso/9780190845711.003.0006

Shields, Christopher 2007. Aristotle on Action: The Peculiar Motion of Aristotelian Souls. *Aristotelian Society Supplementary Volume* 6. 81/1. 139–161. https://doi.org/10.1111/j.1467-8349.2007.00154.x

Shields, Christopher 2014. *Aristotle* (2nd ed.). London, Routledge. https://doi.org/10.4324/9781315863580

Stein, Nathanael 2011. Causation and Explanation in Aristotle. *Philosophy Compass*. 6/10. 699–707. https://doi.org/10.1111/j.1747-9991.2011.00436.x

Sylla, Edith Dudley 2001. *Ideo Quasi Mendicare Oportet Intellectum Humanum:* The Role of Theology in John Buridan's Natural Philosophy. In Thijssen – Zupko 2001. 221–245. https://doi.org/10.1163/9789004452893_015

Thijssen, J. M. M. Hans – Jack Zupko 2001. *The Metaphysics and Natural Philosophy of John Buridan*. Leiden, Brill. https://doi.org/10.1163/9789004452893

Urvoy, Dominique 1991. *Ibn Rushd (Averroes)*. London, Routledge.

Gyula Klima

Teleology, Intentionality, Naturalism*

This paper argues for the contemporary tenability of a "mentalist, Scholastic-Aristotelian" theory of teleological explanations, *pace* contemporary physicalism/naturalism.[1]

I. THE NEED FOR TELEOLOGY IN VOLUNTARY ACTION

Plato's Socrates in the *Phaedo* points to the inevitable need for teleological considerations in the explanation of human actions in the following way:

> [someone might say] that I sit here because my body is made up of bones and muscles; and the bones, as he would say, are hard and have joints which divide them, and the muscles are elastic, and they cover the bones, which have also a covering or environment of flesh and skin which contains them; and as the bones are lifted at their joints by the contraction or relaxation of the muscles, I am able to bend my limbs, and this is why I am sitting here in a curved posture – that is what he would say, and he would have a similar explanation of my talking to you, which he would attribute to sound, and air, and hearing, and he would assign ten thousand other causes of the same sort, forgetting to mention the true cause, which is, that the Athenians have thought fit to condemn me, and accordingly I have thought it better and more right to remain here and undergo my sentence; for I am inclined to think that these muscles and bones of mine would have gone off long ago to Megara or Boeotia – by the dog

* This paper is (a somewhat "spruced up", but basically unrevised version of) the author's translation of a lecture he delivered in Hungarian at a conference entitled *Action and Social Science*, on June 18, 1993, at ELTE, Budapest, Hungary.

[1] Now that logical positivism is a thing of the past, the pretty well-defined requirement of "physicalism" has given way to the (by all appearances deliberate), much fuzzier desire for "naturalism", as if replacing the Greek word with the Latin should make much difference. In any case, the argument that follows is helpfully supplemented on many points by Haldane 1989.

> they would, if they had been moved only by their own idea of what was best, and if I had not chosen the better and nobler part, instead of playing truant and running away, of enduring any punishment which the state inflicts. There is surely a strange confusion of causes and conditions in all this. It may be said, indeed, that without bones and muscles and the other parts of the body I cannot execute my purposes. But to say that I do as I do because of them, and that this is the way in which mind acts, and not from the choice of the best, is a very careless and idle mode of speaking. (Plato: *Phaedo. 98c-99b, tr. Jowett)*

Indeed, it seems that when we are asking about why some conscious, voluntary human act took place, then in the response to this question it is not enough to refer to the act's material or efficient causes. For the voluntary character of the action means that what determined it was the decision of the agent. But we cannot say that it was the decision of the agent that determined the action if the agent's action was determined by some material and efficient causes distinct from the agent's decision.

Consider the following scenario: Mr. Smith is found at midnight walking on the roof of his house in his pajamas: Why? One possible explanation is that Mr. Smith is a sleepwalker, and it is only his disturbed brain-state that makes him perform the dangerous acrobatics. The important thing is that the explanation in this case ends there: he is a sleepwalker, which means precisely that his action was neither conscious nor voluntary, and thus the funny workings of those bones, sinews and muscles Socrates was talking about provide sufficient explanation for his strange behavior.

However, what if we know that Mr. Smith was perfectly conscious and in perfect possession of his power *to choose* to be on the roof? In that case, even if his movements are exactly the same as in the previous case, we would want to know more: Now that we know that his behavior is determined by his choice, we would also want to know why on earth would he *choose to* walk on the roof at the dead of night in his pajamas?

When we learn that he just saw Mr. Taylor slip out from his wife's bedroom window trying to escape through the roof, then we know the explanation of Mr. Smith's strange behavior: he is on the roof *in order to* catch Mr. Taylor, the seducer!

So, now we know why teleological explanations are necessary, at least in the case of voluntary actions: since in their case, what determines the action is *the choice*, but what determines the *choice* is the *intended end*: we can only know why such an action is performed, when we know *to what end* it is performed.[2]

[2] Cf. Aquinas, *Summa Theologiae* I–II, q. 1, a. 1.

II. THE NEED FOR TELEOLOGY IN NON-VOLUNTARY, INSTRUMENTAL ACTION

But is it only voluntary actions that require teleological explanations? In his *On the Principles of Nature* Aquinas replies 'no' to this question:

> We have to know, however, that every agent, natural as well as voluntary, intends some end. But from this it does not follow that all agents recognize this end or deliberate about the end. For to recognize the end is necessary only for those agents whose acts are not determined, but which can have alternatives for [their] action, namely, voluntary agents, who have to recognize their ends by which they determine their actions. However, the actions of natural agents are determined, so it is not necessary that they elect the means to an end. (Thomas Aquinas: *On the Principles of Nature, c. 3, in Klima 2007. 161.)*

So, although in the case of voluntary agents it seems to be obvious that we need teleology for a satisfactory explanation of their actions, since their voluntary actions are determined by their choice, which in turn is determined by the intended end they want to achieve by their action, in the case of non-voluntary agents or even non-voluntary acts of voluntary agents, it might seem that teleology has no place in their explanation. After all, if Mr. Smith was driven to the roof by his somnambulism, then there is no point in asking to what end, or for what purpose he was walking on the roof in the dead of night in his pajamas.

However, as Thomas points out, there is still some point in talking about the end or purpose of some act, even if it is not the agent's own, deliberately targeted, consciously recognized end. For instance, if someone doesn't understand why a circuit breaker tripped in the house, and receives the explanation that it was because there was a power overload in that circuit, then she may still meaningfully ask: alright, but why is the circuit breaker's tripping a good thing, to what end does it trip when there is an overload; after all, what's the point of shutting down the power in the whole house, why is that a good thing if there is no power in the house at all? We should notice that in this case the questioner understands fully well that the action of the circuit breaker is determined to it: if there is an overload, given its mechanism, it trips. And thus, she also knows that the action in question is not voluntary, not determined by the agent's free choice to achieve some end. Yet, the question implies that the action took place for the sake of some end, while it does not imply that the end is the agent's own intended and consciously recognized end. For the questioner perfectly understands that recognizing and intending some end belonged to the principal agent, in our case the electrician, who used the circuit breaker as a means to achieve his end, which in our case is the prevention of an electrical fire that would result from an overload if the circuit were not shut down, and it is the identification of this

end that would provide a satisfactory answer to her question. After all, having a temporary blackout is better than having your house on fire.

Thus, it makes perfect sense to talk about the ends of non-voluntary acts or other phenomena when they are instrumental to achieve the intended end of a principal, voluntary agent. That this is so can be seen even more clearly when we are considering non-voluntary acts of voluntary, human agents, provided they are instrumental to the intended end of another voluntary agent.

For instance, if the by now much-molested Mr. Smith is driven to the roof (in the dead of the night in his pajamas!) neither by his unconscious somnambulance nor by his conscious desire to catch Mr. Taylor, but under the influence of some hypnotic suggestions, then it is, again, entirely pointless to inquire into his purposes, given that his poor, spell-bound bones, and muscles are simply obeying the hypnotic suggestions. But even in this case it is perfectly legitimate, indeed, necessary to question the hypnotizer's purposes, which he will even have to account for in the investigation following Mr. Smith's tragic fall.

Therefore, we can meaningfully talk not only about the end of a non-voluntary act or other phenomenon, insofar as it is instrumental to the intended end of a principal, voluntary agent, but we would not even understand the instrumentality of the phenomenon in question until we know this intended end: we may know what caused the tripping of the circuit breaker (overload), or Mr. Smith's unconscious walk on the roof (hypnosis), but we may still need to know to what end these things happened (avoiding an electric fire, or to murder Mr. Smith and get away with it).

III. A "MINIMALIST THEORY" OF TELEOLOGICAL EXPLANATIONS

So, to sum it all up, based on the foregoing considerations we can provide the following rather plausible, indeed, trivial theory of teleological explanations.

1. One cannot provide a satisfactory explanation of a deliberate, voluntary act without referencing the end, goal, or purpose for the sake of which the agent chose to perform the act in question.
2. Non-voluntary actions and other phenomena may also be goal-directed or purposive, insofar as they are instrumental to the intended end of a principal, voluntary agent.
3. The purposiveness of instrumentally purposive agents can satisfactorily be explained only with reference to the end intended by the principal agent.

For all its plausibility, nevertheless, this simple theory can be attacked in many ways from several angles. In what follows I want to argue through the analysis of only one typical objection for the claim that this objection can appear to be plau-

sible only on the basis of certain metaphysical presumptions that are radically different from those of the Aristotelian metaphysical tradition to which a theory of teleological explanations originally belonged. Indeed, the analysis will show that once these different metaphysical presumptions are explicated, then, for all their "modern scientific character", the objection would rather speak for some general Aristotelian positions.

IV. THE NATURALIST OBJECTION

As can be seen, the theory sketched here is unambiguously "mentalist", insofar as it ties teleological explanations to mental phenomena, conscious recognition of the end (at least in the case of principal, voluntary agents), and free choice or decision to act for the sake of the preconceived end. However, for this reason, to a naturalist committed to explaining away even such apparently obviously goal-directed phenomena in terms of ordinary physical causation obeying the laws of physics, the theory may appear to be committed to a highly suspicious, obsolete metaphysics on many counts.

For according to the first thesis of the theory, what determines the choice of a voluntary agent to act is *the preconceived end* for the sake of which the agent choses to act. However, the end is either (1) the product of the action or (2) some imagination or mental representation of this product. But apparently in both cases we end up with some absurdity. For in the first case, (1) we would have to assume some weird case of "backward causation" wherein a temporally later state of affairs would determine an earlier one. In the second case, (2) the mental representation of the end is either (2a) some purely mental object, an *ens rationis* or figment of imagination, or (2b) some purely spiritual act or (2c) some bodily state of the agent. However, attributing causality (2a) to mere beings of reason or figments of imagination is just as much of a category mistake as attributing colors to numbers. (After all, this is why a scared child can be assured that the boogeyman could not harm him, since the boogeyman is just a figment of imagination.) On the other hand, (2b) a purely spiritual act is already an ontologically dubious item in itself, whatever that is, but it is certainly doubtful whether and how such an item could act on material beings; after all, no-one has yet discovered the formula for mass-spiritual-energy equivalence. Finally, if the mental representation in question is (2c) a bodily state, then there is no good reason for treating it in a special theory burdened with all sorts of obsolete, mystical connotations; on the contrary, what we should do is get rid of these outdated ideas and deal with the relevant phenomena in the framework of our well-working theories of contemporary physical sciences. So, even if teleological explanations may appear to be necessary within common, everyday conversations littered with all sorts of ancient superstitions

("keeping body and soul together", "bless you!", "cross my heart and hope to die", etc.), in their original mentalist form they have no place in the modern, scientific world view. [3]

V. TELEOLOGY AND INTENTIONALITY

In connection with this objection, we should first point out that the causality of the end is of course essentially different from the causality of efficient causes; if it were otherwise, final causes would not constitute a separate genus of the four genera of Aristotelian causes.[4] Thus, the fact that an efficient cause has to be actual in order to exert its causality implies nothing concerning the causality of an end, and so it may be perfectly possible that an end as such does not need to be actual when it exerts its specific kind of causality.

But still, what is that "end as such", and how can it "work" (oh, well, "in its specific way") if it is not yet actual? Well, if we are considering the end of a voluntary, goal-conscious, *successful* act in itself, then it is simply some result of the action. However, this result was not the end or goal of the action only because it resulted from the agent's action, for in this way just any result of the agent's action could be regarded as the action's goal, but that is absurd. For example, if I pour some water into my glass, as a result of this action the air is replaced with water in the glass, but still this replacement of the air is not the goal of my action if I simply pour the water in order to drink it. But this replacement will at once be the goal of my action, if that is what I intend to achieve by the action, say, in the course of an experiment in a chemistry class. What makes a result of the action of a voluntary agent its goal, therefore, is that it is the intended, *intentional object* of the decision of the agent.

But it is precisely at this point that the objector will lose his patience and charge us with the well-known accusation of multiplying all sorts of "weird, mystical entities". So, let us see, once again, whether it is possible to provide satisfactory explanations for voluntary acts without any reference to these "mystical" entities, and if not, whether they are indeed so "mystical" that they could not possibly have a place in contemporary science.

First, let us not forget the reason for introducing these "intentional objects" in the discussion of the foregoing example. As the example shows, the end or goal of an action is distinguished from any other result of it by the intention of the agent, namely, that the agent precisely intended to achieve *this* end by the action and not any other coincidental, perhaps, even necessarily co-occurring

[3] See Nagel 1977. 261–301; Wright 1976; Bedeau 1991. 647–655; Matthen 1991. 656–657; for further naturalistic reductionist attempts, see W. Lycan 1990.

[4] For the issue of the causality of intentions vs. real forms, see Klima 2021.

result of it. Thus, it would be impossible to distinguish between any old result of the action from its goal, if some of its results were not distinguished as the end intended by the agent, as the intentional object for the sake of which the agent performs the action. But this means that we could give up the distinction between result vs. intentional object only if we were also willing to give up the distinction between any old result and the goal or intended end of the action. However, since the action is voluntary precisely because it is performed for the sake of the end voluntarily chosen by the agent, the conflation of the end with any old result would eliminate the distinction between voluntary and non-voluntary acts, thereby invalidating any moral and legal discourse that ties moral and legal responsibility for the act to the voluntary character of the act in question (if your dog bites a passerby, it's not the dog's moral or legal responsibility but yours for not keeping it on a leash). Thus, if we do not want to give up on the meaningfulness of legal and moral discourse, (and why should we?) in the name of some narrow-minded, blinkered, mechanistic conception of science,[5] but still, we do not want to give up on science, indeed, on the possibility of meaningful discourse in general, then we have to work with a scientific conceptual framework that can accommodate discourse about intentional objects as well. But it seems that this is prevented precisely by the weird, mystical character of intentional objects. However, we also know that there used to be a certain kind of science, namely, scholastic Aristotelian science, in which physics and ethics, efficient causality and teleology, physical and intentional objects each found their place in perfect harmony. Let us see, therefore, whether those intentional objects are indeed as insufferably mystical or perhaps it is only the historically understandable, but conceptually rather contingent anti-Aristotelianism of early modern science that wraps it into an only to us impenetrable mystical mist.

VI. INTENTIONALITY VS. PHYSICALISM

As we can see, the intentional object of an act and its result often coincide, namely, when the act is successful, and realizes the agent's intended end. In this case, therefore, there is nothing mystical in the intentional object, because it coincides with an ordinary physical object, the result of the action. However,

[5] Since one of the referees of the original version of this paper felt somewhat offended on behalf of naturalists by this characterization (I quote: "which suggests as if those who do not acknowledge the necessity of teleological explanations were somehow not epistemic and cognitive peers of the author and get their drive from misunderstandings and superstitions (e.g., fear of "weird, mystical entities")"), I should perhaps clarify that I am not lumping together all naturalists under this phrase: I am only talking about those who do fit under it; examples would be especially from the 18th century, although we could easily find later examples as well, but when it comes to hurt feelings, *nomina sunt odiosa*.

even in this case, there is a difference between their conditions of identity. This can be seen most clearly if we consider the fact that one and the same result of a voluntary act can both be and not be the intentional object of a certain result of a voluntary act. As is well-known, Oedipus wanted to marry Iocaste, but he did not want to marry his mother, although, as a matter of fact, Iocaste was his mother; ergo, as a result of their wedding, Oedipus married his mother. In this case, the intentional object of Oedipus' voluntary act, namely, to have Iocaste as his wife, coincides with the result of his action, namely, that he married his mother and thus he had his mother as his wife. And yet, this result was definitely not the intentional object of his act, indeed, he wanted to avoid this result throughout his life. Of course, what accounts for the difference is the fact that Oedipus did not know that Iocaste was his mother, and thus he did not know that marrying Iocaste was the same as marrying his mother. Had he known her identity, he could not have wanted the one without the other. Thus, it is part and parcel of the conditions of identity of the intentional object how this object is represented in the voluntary agent's mind, or in scholastic terminology, what is that *ratio* under which the agent's mind represents it.[6] However, having identified this *ratio*, we can identify the intentional object as well without further ado, and thus it can be applied in an exact fashion in both the scientific and the moral description of Oedipus' behavior.

To be sure, the above-described imaginary champion of modern science will probably not be any happier with this "solution", in which now instead of one mysterious entity he has to deal with two: the intentional object and its *ratio*, not to mention the obsolete, barbaric terminology.

Well, in the age of "nice quarks", we may perhaps set aside the Johnny-come-lately humanist squeamishness about the terminology; so, we may focus on the things themselves no matter what we call them. In any case, the objector can still say that since we ended up with the result that the goal/end/purpose of some voluntary agent is characterizable as such on account of how some result of the action is represented in the agent's mind, we can get around the entire hocus-pocus by focusing on this mental representation itself, which we can then describe as some ordinary neurophysiological phenomenon in terms of a successful physicalist reduction. In this way, we can of course still keep the language of teleological explanations, perhaps, for some practical, or nostalgic, or maybe historically important reasons, while always knowing that this simple,

[6] For a scholastically inspired formal treatment of a logically similar intentional paradox, see Essay 5 of Klima 1988. For the relevant notion of *ratio*, basically, the intelligible content of an object grasped by a mental representation that determines the identity conditions of the mental representation itself, see Klima 1993; 2015. Also note that with this understanding, the phrase can also refer to the mental representation itself, strictly identified in terms of its semantic content, regardless of what encodes this content in some or another particular (type of) medium. So, the "multiple realizability" of a *ratio* is *ab ovo* built into its notion.

obsolete language just stands in for a more complex, but scientifically reliable physicalistic, neurophysiological explanation (much like ordinary loose talk about hot and cold stands in for the scientifically exact notion of temperature, analyzable in terms of the mean kinetic energy of particle movement).

In connection with this reasoning, there are two points that are clear at once. On one hand, its validity is at least highly dubious until someone actually carries out the requisite "physicalist reduction", for until then the "neurophysiological phenomena" in question have no more explanatory power than the *rationes* of the scholastics, whatever those are. On the other hand, it should also be clear that on account of a possible successful physicalist reduction all moral and legal discourse about responsibility will have to be reduced just as the psychological discourse about goals and voluntary choices grounding it. It is a good question, then, whether the need to get rid of "mystical entities" can justify such a program, especially if it turns out that despite their absence from modern (physicalist) scientific discourse they are not that mysterious after all. And it is yet another question whether such a reduction can be carried out at all, whether it would not run into some principled, conceptual obstacles that would render the task impossible to complete. Let us look at this last worry first.

As has been seen, the distinction between physical objects and intentional objects was prompted by the difference between the intended end and any other physical result of some voluntary act. We could also see that intentional objects had to be distinguished from physical objects, even in cases when they actually coincide with physical objects, because the intentional objects have different criteria of identity (see marrying Iocaste vs. marrying Oedipus' mother). There is nothing surprising in this. We know that the criteria of identity of things can vary with the ways we refer to them, since these ways determine their classification, distinction, counting, and re-identification. For instance, to the question of how many things there are in this room, the answers may range from the number of macroscopic substances to the number of their macroscopic or microscopic, even subatomic parts, their attributes, various collections, relations, or the number of facts, events, or processes taking place, not to mention the number of concepts or thoughts we are engaging right now. Thus, in connection with the reduction program, the question is whether the neurophysiological phenomena in question are such that their criteria of re-identification are at least as good for the identification of intentional objects as are the *rationes* of the scholastics, whatever the ontological status of the latter.

Now it is clear that the *rationes*, as mental representations, are identifiable by means of linguistic expressions, although their criteria of identity are not the same as those of linguistic expressions. These mental representations are conceptual structures that can be expressed in terms of radically different linguistic structures in different languages. Think, for instance, of the different syntactical structures by which a negation, which is certainly a distinctive element of

a conceptual structure, can be expressed in different languages. But then the same conceptual structure will correspond to different linguistic representations in different languages, and so, also, there will be different neurophysiological phenomena taking place in the nervous systems of the speakers of different languages while they are processing the same conceptual structures in their respective languages. Indeed, in the nervous system of the same bilingual speaker there will be different neurophysiological phenomena taking place while the speaker is processing the same conceptual structure in different languages (as in preparing a translation). Thus, the criteria of identity of neurophysiological phenomena are always different from those of the *rationes*, the conceptual structures identifying voluntary agents' intended ends as their intentional objects, and so they will never provide a good means for the identification of these intentional objects, and thus for their elimination in a physicalist reduction, while keeping teleological discourse meaningful.

At this point someone might object, of course, that this piece of reasoning does not prove that the same conceptual structures represented by different linguistic structures must correspond to different neurophysiological phenomena, as it is possible that on a deeper neural level the processing of different linguistic structures is mapped onto some neurophysiological phenomena directly matching the conceptual structures in question.

The objection may seem to be legitimate, but it does not really help the completion of the physicalist project. In the first place, we just don't know whether there are some such "deeper" neurophysiological phenomena.[7] And even if there were, they would not be much help. For the mere possibility of a one-one match between conceptual structures and some deeper neurophysiological phenomena is not sufficient for the viability of the physicalist project, because for the viability of the project it is necessary that this one-one match is necessary, and not merely contingent, for it is only this condition that guarantees that any possible conceptual structure is unambiguously matched with the correct neurophysiological phenomenon. For it is only this necessary connection that can guarantee that the description of any possible conceptual structure is correctly eliminable in terms of the corresponding neurophysiological description, much in the same way as if there were no unambiguous machine code translation of the instructions of a high-level programming language, then the latter could not

[7] To be sure, if there were such "deep" neurophysiological phenomena, they would constitute a uniform mental language, a "language of thought", à la Jerry Fodor, for all humans, encoded in those phenomena. I raised several doubts concerning there being such a uniform "language of thought", regardless of whether it is encoded in a material or some "spiritual" medium, and also concerning the theoretical usefulness of positing such a uniform mental representational system here: Klima 2012. Of course, I'm not alone with such doubts, but perhaps my arguments present a rather different perspective from usual criticisms.

in principle be eliminable in the machine code, that is to say, a possible correctly written program would not compile.

However, the previous piece of reasoning showed precisely that there is possibly no one-one correspondence between conceptual structures and neurophysiological phenomena; therefore, the logically necessary connection between the two required by the viability of the physicalist project does not obtain. For the point of the argument is that neurophysiological phenomena as such, even if they may correlate with conceptual structures, are essentially differently classifiable and identifiable from conceptual structures. And this is so because conceptual structures as such are essentially representative; thus, it belongs to their conditions of identity what and how they represent. However, from the study and description of a neurophysiological phenomenon in itself it will never be apparent what and how it represents, as it only logically contingently correlates with its object, whereas a conceptual structure is identifiable precisely on the basis of what and how it represents.[8] So, studying the neurophysiological phenomena can give us no more information about their semantic, representational features than looking into the magnetic polarities of a computer hard drive would yield its contents without knowing the code that establishes by the logical necessity of conventional encoding the connection between magnetic patterns and what they represent under that code.

To be sure, this argument is not to be read as an attempted knock-down proof against the possibility of the logically contingent (but perhaps causally necessary) identity (or just correlation) of concepts *(rationes)*, or even all sorts of "mental states", and neurophysiological phenomena.[9] In the first place, there is nothing wrong with the idea, especially in the case of sensory states or processes (such as acts of perception, sensory memory, or imagination, etc.), which obviously require for their occurrence (or may even consist in) the activity of some (external or internal) sense organs. Perhaps, in the case of higher intellective functions we may have good reasons to doubt the possibility of such identifications,[10] but that is not the point of the foregoing argument. The argument rather intends to show

[8] On the necessity of the connection between object and concept, see Klima 1991. The basic idea in a nutshell is that a concept is nothing but the form of the object in the mind. Since the concept is the form of the object, its reception in the subject necessarily makes the subject actual in regard of the form of the object, although not in the same way as it makes the object actual: for it makes the object *to be actual* in its real being in regard of the form, whereas it does not make the mind actual as it makes the object actual in its real being, but it makes the mind *to be actually cognizant* of this form. Thus, the formal content of the actuality is the same, but the mode of actuality is different, which stems from the natural difference of the recipients. As Thomas states in many places: *receptum est in recipiente secundum modum recipientis*.

[9] I am grateful to another anonymous referee for providing me with the opportunity for this clarification by making the objection that my argument does not prove that this contingent identity or just correlation is not possible, "or, at any rate, no argument is given in the paper that the viability of the physicalist project is ruined by such a contingent association."

[10] For my latest musings on Aquinas' relevant argument and its implications for AI, see Klima 2022.

that even if all mental representations were in fact logically merely contingently identical with neurophysiological phenomena, that logical contingency would in principle prevent carrying out the physicalist/naturalist reduction of teleological explanations of voluntary actions, because we just could not have a reliable code allowing us to read off intentions from brain scans, which any such reduction deserving the name would have to be able to carry out.

VII. NATURALISM VS. MODERN ARISTOTELIANISM

Neurophysiological phenomena as such, therefore, cannot replace the conceptual structures that play a crucial role in identifying the intended ends of voluntary agents. But what are these conceptual structures if they cannot be identified either with linguistic or with neurophysiological phenomena? This is what makes them so disturbing, namely, that apparently they always slip out from the sphere of "reliable" entities that are clearly identifiable by means of scientific methods!

But where is it set in stone, we may ask, that it is only the scientifically identifiable entities that are "reliable"? If we are looking for scientific exactitude, we shall sooner find it in mathematics than in physics, and if there is anything laudable in the glorious 20th century, then it is the fact that it is the thinkers of that century that made the mathematical modelling of conceptual structures possible. Thus, if within this mathematical framework we are able to obtain an exact way of grasping these conceptual structures, the scholastic *rationes*, then their "mysticism" as well as that of their intentional objects will be just as problematic as mathematical entities are, that is, from the point of view of a scientific world view, not a whit. To be sure, this does not mean that the ontological status of mathematical entities or conceptual structures and their intentional objects would not pose a genuine philosophical problem. But that is the philosophical problem we cannot and need not go into at this point.[11] After all, the issue here is not the metaphysics of intentions, but the irreducibility of teleological explanations and the possibility of their integration into a more broad-minded scientific project, closer to the scientific ideals of scholastic Aristotelianism than to the ideals of a Newtonian-Laplacian mechanics. At any rate, it should be clear that it was precisely the naturalistic objection that prompted the philosophical considerations that in turn directly led to this traditional philosophical problem, thereby pointing us toward a philosophical, Aristotelian ideal of science, instead of pointing toward a narrow-minded, blinkered physicalism,[12] chasing in vain the pipe dreams, or at least so far only the promissory notes, of a physicalist reduction.

[11] See, nevertheless, Klima 2014; 2015.

[12] Which phrase, again, is not meant to derogatively apply to all possible and actual forms of naturalism.

VIII. CONCLUSION: WHY *MODERN* ARISTOTELIANISM?

All this should not mean, though, that instead of the theory of relativity now we should study Aristotle's *Physics* (although we should not neglect that either) or that we should put all our faith into the Aristotelian, rather the Mendelevian elements. On the contrary, we should rather strive to present such a philosophical understanding of the theory of relativity and of the Mendelevian elements, and all the rest of modern scientific facts, theories, and phenomena that, just like the Aristotelian tradition, would not render human discourse in the humanities meaningless.

However, we should also notice that besides setting up some loose analogy and a vague value-requirement, the Aristotelian philosophical tradition can provide us with some more direct help. As the foregoing analysis of the problem of teleological explanations illustrated, the conceptual framework of the Aristotelian tradition, although by and large may be "out of fashion", in a modern formal interpretation can not only live up to the modern requirements of scientific exactitude, but it can even fill in its philosophical gaps. In particular, it shows that teleological explanations can function as perfectly legitimate scientific explanations, once we understand their specific character, and we do not try to squeeze them into the straitjacket of some unfounded scientific ideal in terms of a physicalist, naturalist reduction. But, having seen this much, we could also understand how we can provide teleological explanations even for the agency of non-voluntary agents, insofar as it is instrumental in the agency of voluntary agents. As we can also see, such non-voluntary, instrumental agency can perfectly be explained without any reference to their end, for it is only their instrumentality that would be inexplicable without reference to the intended end of the principal agent. Thus, non-voluntary natural phenomena can perfectly be accounted for in terms of some physicalist explanation, seeking to understand merely what accounts for the coming to be or sustaining of such phenomena.

However, such explanations will never satisfy someone for whom the whole of nature and any and all phenomena in it are instrumental to some overarching intelligent purpose. As we have seen, the possibility of "complete" naturalistic explanations will never eliminate the legitimacy of teleological questions, and thus the entire modern army of the "Newtons of a blade of grass"[13] will not eliminate the eternal human question: "And for what purpose is the whole creation?"[14]

[13] Cf. "we may boldly state that it is absurd for human beings even to attempt it, or to hope that perhaps someday another Newton might arise who would explain to us, in terms of natural laws unordered by any intention, how even a mere blade of grass is produced" (Immanuel Kant 1790/1987. Part II, sect. 75, n. 400. 282.)

[14] Imre Madách: *The Tragedy of Man*, tr. Tomschey, O., Budapest, Madách Irodalmi Társaság. 2000. 4. l. 97. For the scholastic idea of *natural teleology* understood in terms of an

REFERENCES

Aquinas, Thomas 1250/2007. *On the Principles of Nature*. In Klima (Ed.) *Medieval Philosophy: Essential Readings with Commentary*. Oxford, Blackwell Publishers.

Bedeau, Mark 1991. Can Biological Teleology Be Naturalized? *The Journal of Philosophy*. 88. 647–655. https://doi.org/10.5840/jphil1991881111

Haldane, John J. 1989. Naturalism and the Problem of Intentionality. *Inquiry*. 32/3. 305–322. https://doi.org/10.1080/00201748908602196

Immanuel Kant 1790/1987. *Critique of Judgment*. Trans. Werner S. Pluhar. Indianapolis, Hackett Publishing Company.

Klima, Gyula 1988. *Ars Artium: Essays in Philosophical Semantics, Mediaeval and Modern*. Budapest, Hungarian Academy of Sciences.

Klima, Gyula 1991. Ontological Alternatives vs. Alternative Semantics in Medieval Philosophy. *S: European Journal for Semiotic Studies*. 3-4. 587–618.

Klima, Gyula 1993. The Changing Role of *Entia Rationis* in Medieval Philosophy: A Comparative Study with a Reconstruction. *Synthese*. 96. 25–59

Klima, Gyula 2012. Ontological Reduction by Logical Analysis and the Primitive Vocabulary of Mentalese. *American Catholic Philosophical Quarterly*. 86. 303–414.

Klima, Gyula 2014. The Problem of Universals and the Subject Matter of Logic. In Penelope Rush (Ed.) *The Metaphysics of Logic*. Cambridge, Cambridge University Press. 160–177.

Klima, Gyula 2015. Mental Representations and Concepts in Medieval Philosophy. In Klima Gyula (Ed.) *Intentionality, Cognition and Mental Representation in Medieval Philosophy*. New York, Fordham University Press. 323–337.

Klima, Gyula 2021. Form, Intention, Information: From Scholastic Logic to Artificial Intelligence. In Ludger Jansen – Petter Sandstad (Eds.) *Neo-Aristotelian Perspectives on Formal Causation*. London, Routledge. 19–39.

Klima, Gyula 2022. Language and Intelligence, Artificial vs. Natural, or What Can and What Cannot AI Do with NL? In Henning Bordihn – Géza Horváth – György Vaszil (Eds.) *Proceedings of the 12th International Workshop on Non-Classical Models of Automata and Applications* (EPTCS 367). https://doi.org/10.4204/EPTCS.367.1

Lycan, William. G. (Ed.) 1990. *Mind and Cognition*: *A Reader*. Cambridge, Blackwell.

Matthen, Mohan 1991. Naturalism and Teleology. *The Journal of Philosophy*. 88. 656–657. https://doi.org/10.5840/jphil1991881112

Nagel, Ernest 1977. Teleology Revisited. *The Journal of Philosophy*. 74. 261–301.

Wright, Larry 1976. *Teleological Explanations*. Berkeley, University of California Press.

intrinsic drive in all creatures to share in divine perfection to the capacity of their nature by completing the possible perfections of their nature, each in its own way as God intended, see, e.g., Aquinas, *Contra Gentiles*, lib. 3 cap. 2; *Summa Theologiae* I, q. 44 a. 4 co.

Dániel Kodaj

The Metaphysics of Spooky Teleology*

Teleology is dead – it was killed by modern science. Indeed, its demise was already announced by the founders of modern philosophy. "The whole category of causes that people are in the habit of seeking by considering the purposes of things is of no use in the study of physics," Descartes wrote (1641/2008. 40), while Bacon famously opined that "inquiry into final causes is sterile, and, like a virgin consecrated to God, produces nothing" (*De Augmentis Scientiarum* bk. iii ch. 5; quoted by Woodfield 1976. 3). Hobbes likewise ridiculed explanation from final causes: "If you desire to know why some kind of bodies sink naturally downwards toward the earth, and others go naturally from it; the Schools will tell you out of Aristotle, that [...] the cause why things sink downward, is an endeavour to be below: [...] as if stones and metals had a desire, or could discern the place they would be at, as man does; or loved rest, as man does not" (Hobbes 1651/1998. 450f). The idea of robust goal-directedness ("spooky teleology",[1] as I'll call it) disappeared from our conception of nature. Science and philosophy got rid of it once and for all. It may enjoy a vestigial presence in folk metaphysics (Kelemen et al. 2013, Rose–Schaffer 2017), while some Thomists, far away from the philosophical mainstream, desperately hold onto an obscure Aristotelian version of it, but teleology is really just a relic of the past. Or so we are told.

But what is it, exactly, that modern science exorcised? Since robust teleology is an extremely unfashionable topic, it is virtually never discussed in contemporary analytic metaphysics. It is missing from the metaphysician's conceptual toolkit, even as a logical possibility. My goal is to put it back there.

* Thanks to audiences at the Central European University, ELTE, and the 2023 New Generation Research Exchange conference at Zagreb for comments on earlier versions of the paper, and for the anonymous reviewers of this journal. I am especially grateful to László Bernáth, Gergely Kertész, and Tamás Paár for conversations on this topic.

[1] "Spooky" is term of art in analytic metaphysics for posits that fly in the face of materialism (see e.g. Dupré 2012 for such a use of the word). The term probably originates in Einstein's description of quantum entanglement as "spooky action at a distance" (*spukhafte Fernwirkung*, Einstein et al. 1971. 158).

This is an exercise in conceptual engineering, not an attempt at historical interpretation. I would like to find a definition of spooky teleology that satisfies the following desiderata: (i) it is built from concepts that contemporary metaphysicians understand (it is ***idiomatic***), (ii) it captures goal-directedness (it is ***adequate***), and (iii) it is ***spooky*** in the sense that mainstream physicalists are likely to deny that anything answers to it in reality. In Section 1, I present a handful of potential definitions culled from the literature, and in Section 2, I develop my own proposal which is based on Braithwaite's (1947) concept of plasticity.

This is what I'm selling, nothing more and nothing less. Why should anyone buy it? I believe that a clear understanding of spooky teleology is important for three reasons.

First and most important reason: You cannot disagree with something that you don't understand. If you think that modern science killed spooky teleology, you should have a reasonably clear idea of what spooky teleology would be if it existed. Otherwise you are not entitled to assert its demise.

Let me offer a quick case study to drive this point home. A recent paper about teaching methodology draws attention to the fact that biology students instinctively think of functions as the causes of functional traits. Students display a bias for what the authors of the paper call "ontological teleology" – an allegedly unscientific notion that, in the authors' view, should be carefully eradicated. Here is how they explain the concept in question:

> Ontological teleology assumes that an explanandum came into existence because of its function within the organism or ecosystem. Some instances of ontological teleology do not specify how exactly the formation of the explanandum became directed towards the function, but other instances of ontological teleology attribute the striving towards function to the intention of a force that sets functionality as a goal. (Trommler–Hammann 2020. 4)

I submit that this description is so vague that it does not describe an intelligible conception of spooky teleology. At certain points (e.g. when it mentions forces that have intentions), the text borders on the nonsensical. Of course, it is possible that the text accurately reflects the way students think, because students are simply confused. But even in that case, biology education would be better served by a rational reconstruction of unscientific intuitions. A clear picture of spooky teleology would allow instructors to explain why we are justified to think that it is absent from nature.

Further, a clear picture of spooky teleology can be useful for interpreting historical doctrines. Although my goal here is explicitly ahistorical (all I want is an ***idiomatic***, ***adequate***, and ***spooky*** conception), I think that a good definition can help us make more sense of premodern ontologies and compare them to the

antinaturalist approaches that proliferate in contemporary analytic metaphysics (see e.g. Koons–Bealer 2010).

Last but not least, a workable definition of spooky teleology is important for reassessing the received view about the death of teleology at the hands of modern science. It is conceivable that a clear understanding of spooky teleology will deliver a more nuanced conclusion. At the end of the paper, I will suggest that spooky teleology could very well be real even in light of modern science.

I. IN SEARCH OF SPOOKY TELEOLOGY: THE STATE OF THE ART

This section will present increasingly complex approaches to spooky teleology, culled from contemporary metaphysics and from discussions about earlier versions of this paper. I cannot guarantee that there are no further candidates in logical space (or even in print), but I tried to be as comprehensive as I could be.

Let me briefly mention, and set aside, an obvious candidate for spooky teleology: divine providential activity. If our world is structured in such a way that it realizes the goals of a divine being, then, clearly, the world contains spooky teleology in some sense. But if this is the whole story about goals in nature, then we don't really have a distinctive conception of spooky *teleology*. What is antithetical to contemporary naturalism, in the picture that we are considering, is the presence of a divine being who fiddles with the layout of the universe. Teleology is not a robust additional ontological component in this account, but something that supervenes on the divine will. Another way to put this point is that teleology is extrinsic to created beings if it is wholly grounded in divine providential activity. In contrast, the kind of spooky teleology that this paper seeks to understand is an *intrinsic* feature of things, not something imposed on them from outside.[2] I will treat the intrinsicality requirement as an implicit component of the criterion of ***adequacy*** in what follows.

[2] One of the reviewers raises the objection that if divine teleology is completely extrinsic, then "it is also 'completely extrinsic' to a TV set that it is for watching TV programs, or in general to any tool that is in the service of human intentions. If this sort of instrumental teleology is excluded from the picture altogether, then one wonders what can remain for the teleology of non-rational or even generally non-cognitive agents, apart from mere intrinsic spookiness without a purpose. Furthermore, if created natures are created so that their inherent mechanisms serve some divine purpose (such as the perfection of the universe), then why would this inherent drive toward a divine purpose be extrinsic to them?" Teleology is intrinsic if directedness, or the capacity for it, is part of the nature of the entity in question. TV sets are not intrinsically teleological in this sense, while plants could very well be. The fact that directedness has an external causal origin (namely, God's creative activity) or that it involves external objects as means or ends does not make it extrinsic, in my terminology. What I am trying to rule out (what I treat as 'completely extrinsic') is the kind of teleology that consists in God's arranging a collection of mechanistic systems in a pattern that serves His purposes.

1. Mentalism

Perhaps the simplest approach to spooky teleology is the view that I will call *mentalism*. According to this view, teleological systems are literally minded, they exercise their will to realize their goals. This conception is intelligible and very spooky. According to standard contemporary physicalism, seeds are not striving to grow into trees, and foetuses are not consciously trying to become healthy babies, as the mentalist claims if she treats such beings as teleological.

Not only is mentalism idiomatic and spooky, it is also quite familiar, since it is often used to ridicule the idea of teleology (cf. the Hobbes quote in the introduction). However, rhetorical effectiveness aside, mentalism is not ***adequate***, because it is either weird and irrelevant or it is empty.

The thesis that seeds, foetuses etc. are literally minded is too weird to be a good conception of the kind of spooky teleology that Darwinism is supposed to have exorcised. When someone who is not in tune with modern science claims that foetuses have a *telos*, she does not mean that they are formulating plans in a hidden homunculus mind. Note that mentalism of this egregious sort is quite different from panpsychism, so one cannot drag panpsychism into the dialectic to make mentalism look less weird and irrelevant. Panpsychists believe that all physical objects have phenomenal states, while the mentalist believes that plants and foetuses have agential cognitive states. Panpsychists attribute phenomenal consciousness to *all* things, while mentalists attribute cognitive capacities to *some* things that clearly don't have such capacities. So these doctrines are completely different, and the respectability of panpsychism does not transfer to mentalism.

Alternatively, if the mentalist does not claim that seeds, foetuses etc. are literally agents but she claims, instead, that they are *similar* to agents, then mentalism is empty: it is not not a conception of spooky teleology but an invitation to provide one. It is quite obvious that teleological systems resemble conscious agents in some respect. The task is to explain why.

2. Retrocausality

One could think of spooky teleology as a form of retrocausality: a future state (the end or goal) causes the activity that leads to it. There are two obvious advantages of this view. First, it is ***idiomatic***: retrocausality is a familiar topic in contemporary metaphysics. Second, conceiving of teleology as retrocausality endows final states with the same metaphysical significance that causes can in principle have. For example, if causes explain their effects, then the retrocausal conception will entail that goals explain the actions leading to them. If causes make their effects more likely, goals will make the means more likely. And so on.

Unfortunately, the retrocausal conception is ***inadequate***. As Hawthorne and Nolan (2006. 274) point out, nonexistent future events cannot cause anything, hence the retrocausal conception entails that teleological processes necessarily reach their goal.[3] But this is clearly false. The development of a foetus (assuming, for the sake of argument, that it is goal-directed) can be arrested in all sorts of extraneous ways. If a three-month-old foetus dies because of an accident that has nothing to do with its own developmental processes, then it fails to realize its goal, hence its end state cannot cause anything from the future. Yet the foetus's development occurred for the sake of an end. So we have a teleological process that is not retroactively caused by its future end.

A fan of retrocausality could bring in merely possible events and say that arrested teleological processes are caused by states in close possible worlds where the end is realized. But this modification makes the view ***unidiomatic***, because interworld causation is a highly unfamiliar idea. Indeed, it is not only unfamiliar but plausibly impossible: Aristotelians would say that the merely possible cannot act, and even common sense suggests that nonactual things lack causal powers.[4] Moreover, the modified view continues to be ***inadequate***, because it portrays teleology as extrinsic to the actual world in cases where a goal-directed process is arrested. But goal-seeking, whether successful or not, should be intrinsic to things, according to the rules of ***adequacy***. So spooky teleology is not retrocausality.

3. Causal powers

In an ironic twist of history, causal powers re-emerged from the grave that early modern philosophers dug for them and they have gained a considerable following in contemporary metaphysics.[5] Some of their proponents believe that they also help make sense of teleology. Here's how Robert Koons articulates this approach:

> [C]ausal powers are inherently teleological. To have the power to produce *E* in circumstances *C* is to have the *C*-to-*E* transition as one of one's natural functions. Indeed, as George Molnar has pointed out (Molnar 2003), the ontology of causal powers

[3] On a plausible presentist view, no future events (and hence not even realized future ends) exist. Thanks to a reviewer for pointing this out.

[4] As a reviewer remarks: "If the bogeyman is not actually there in the kid's room, then the kid can be reassured that it cannot harm her." Thanks for the additional point and the example.

[5] See Ott (2009) on the early modern demise of causal powers. The history of their resurrection is yet to be written; good overviews of the state of the art include Corry (2019) and McKitrick (2018).

> builds intentionality into the very foundations of natural things. To have a power is to be in a kind of intentional state, one that is in a real sense "about" the effects one is pre-disposed to produce. (Koons 2021. S899)

A slightly different route from powers to teleology is explored by Paolini Paoletti (2021). He constructs a variety of technical concepts (weakly teleological, very weakly teleological, strongly teleological), but for illustrative purposes, the following rough idea will do: *x*'s causal power *P* is teleological iff (a) neither the activation nor the possession of *P* by *x* depends ontologically on any other power of *x* or on categorical facts, and (b) all other powers of *x* depend on *P*.

These conceptions are ***idiomatic***, since causal powers are well-known (although not universally loved) tools in the analytic metaphysician's toolkit nowadays. These conceptions are, however, ***inadequate***. Consider an electron's power to repel negatively charged particles. This property has physical intentionality in Molnar's and Koons' sense, yet the process that it gives rise to – repulsion according to the laws of electrodynamics – is clearly not teleological. The electron does not strive to repel other electrons, it does not generate an electric field for the sake of repelling other electrons, there are no better or worse ways for it to repel other electrons, and so on. Similar remarks apply to Paolini Paoletti's approach, since an electron's negative charge is plausibly seen to fulfil condition (a), while fulfilling condition (b) would not add anything interesting to the picture with respect to goal-directedness.

A fan of causal powers could insist that even a blind physical process like electromagnetic repulsion is teleological – the realm of natural ends is much wider than fans of teleology used to think. I grant that one can use "teleology" in such a way that electromagnetic interactions qualify as teleological (just as one can use "leg" in such a way that dogs qualify as five-legged animals). But this is not the sense of "teleology" that students of spooky teleology are interested in. There is nothing spooky about the interaction of charged particles. Identifying teleology with the directedness of causal powers does not bring us any closer to understanding what spooky teleology would be if it existed. We need a conception that is much more narrow and that obviously clashes with mainstream reductive naturalism (as causal powers do not; cf. Corry 2019).

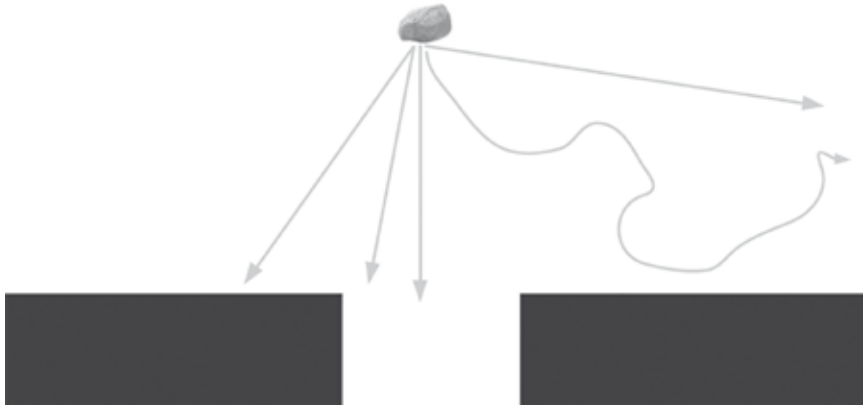
4. The Hawthorne/Nolan Model

In an interesting exercise in conceptual engineering, John Hawthorne and Daniel Nolan (2006) tried to capture a form of characteristically Aristotelian teleology using language reminiscent of the 'least action' principles of modern physics. Their conception is ***idiomatic*** by design. Moreover, since it envisions a

form of causation that (in light of modern science) does not exist, it is sufficiently ***spooky***. I will argue that the definition is nonetheless ***inadequate***.

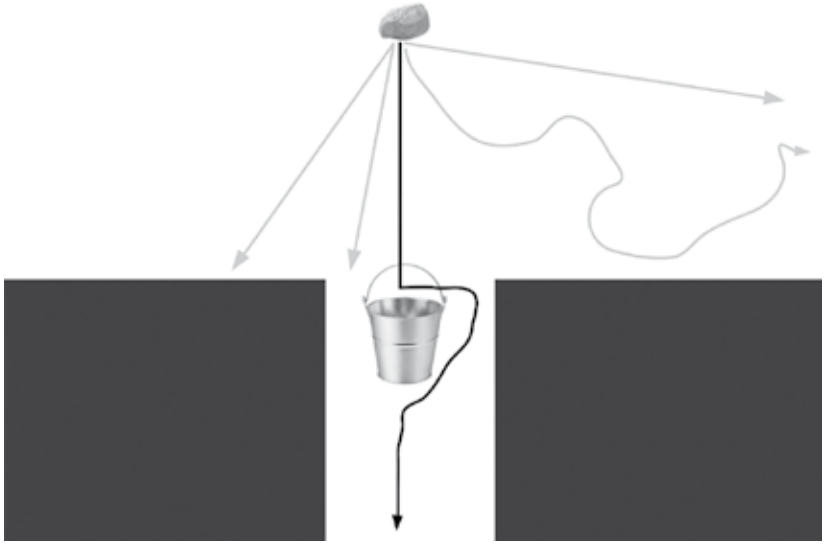
The stock example that Hawthorne and Nolan use is a rock seeking its natural place at the center of the Earth. For the sake of argument, I will neither endorse nor dispute the claim that this is a good paradigm of Aristotelian final causation. As mentioned in the introduction, issues of historical interpretation are orthogonal to the present inquiry.

Suppose that we lift our goal-seeking rock and we release it above a well. Once released, a number of logically possible trajectories are in principle available to the rock:



In the Hawthorne/Nolan model, goal-directedness is defined as the disposition to choose a specific type of trajectory. Very roughly, the idea is that the rock (while seeking its natural place at the center of the Earth) 'chooses' a trajectory that takes it to to its end at least as fast as any other available trajectory. "Choosing" is understood in a non-mentalistic fashion, as a disposition to move in a certain way. The mark of teleology, on this view, is the disposition to choose a path that is optimal in some sense.

This idea must be Chisholmed a bit. According to the rudimentary definition sketched above, the rock should choose a path that takes it to the center of the Earth at least as fast as any other path could. But then the rock should go around the bucket in the following scenario:



Circumnavigating the bucket would help the rock reach its goal quicker, but this is clearly not how rocks move. To avoid this counter-intuitive consequence, Hawthorne and Nolan propose the following sophisticated criterion:

> (HN) At any time *t*, the rock will follow a continuation of its path in such a way that, for some period after *t*, the rock has greater end-velocity [=moves faster toward its goal] than it would have on any alternative path that is compatible with the relevant constraints.

(HN), unlike the earlier rudimentary definition, predicts that the rock will fall into the bucket instead of circumnavigating it, because doing so yields a greater end-velocity in the period when the circumnavigating trajectory proceeds sideways. So (HN) is at least *prima facie* applicable to actual phenomena. The gist of (HN) is that something behaves teleologically iff its path is always locally optimal in the sense that it takes the object toward its end at least as fast as any other available path.

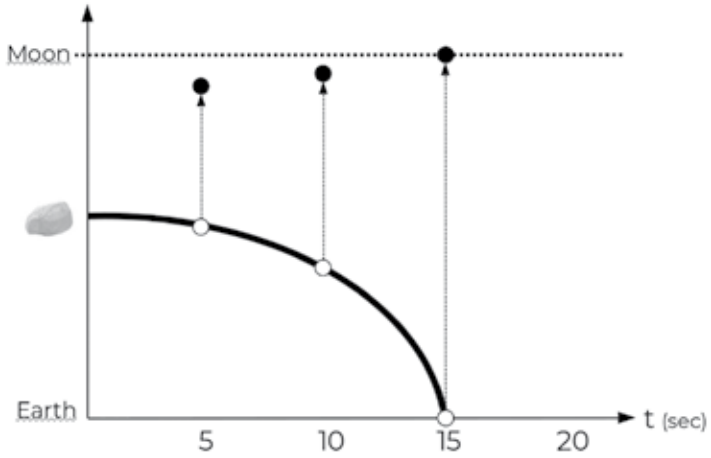
Is this conception ***adequate***? Not really. First of all, it is subject to counter-examples even within the narrow domain of goal-seeking rocks. Consider the Moon Stone, an unusual (merely possible) piece of rock whose behaviour is governed by the following rules:

> (A) If the Moon Stone is not on the surface of the Moon and it is not physically restrained, then it makes an instantaneous jump toward the Moon every five seconds. The size of the jump at *t* is $0.999*\Delta(t)$ if $\Delta(t) > 1$ km

and $\Delta(t)$ otherwise, where $\Delta(t)$ is the rock's distance from the nearest point on the surface of the Moon at t.

(B) When the Moon Stone is not jumping and it is not on the surface of the Moon, it obeys (HN).

If the Moon Stone is released somewhere on Earth, its trajectory looks something like this:



According to the Hawthorne/Nolan model, the Moon Stone is teleologically directed at the center of the Earth between $t = 0$ sec and $t = 15$ sec, since it obeys (HN) at every point in that period. But, clearly, the Moon Stone is *not* teleologically directed at reaching the center of the Earth. Its goal (assuming, for the sake of argument, that it has one) is to get to the Moon. And it works towards this goal with admirable efficiency, much more efficiently than the Earth-bound rock of Hawthorne and Nolan. So the Hawthorne/Nolan model, although ***idiomatic*** and moderately ***spooky***, is ***inadequate***.

One could perhaps get around this problem by more Chisholming, for example, by requiring that the end-velocity be well-defined throughout the trajectory. But then the Moon Stone will qualify as non-teleological, which also sounds wrong.

A reviewer raises the objection that the Moon Stone's behaviour is not only physically impossible but contrary to our sane intuitions about motion. So it hardly qualifies as a counterexample. I reply that nomologically impossible scenarios provide perfectly good counterexamples if (HN) is meant to be a metaphysically necessary criterion of goal-directedness. The Moon Stone is metaphysically possible, it is goal-directed (to the extent that stones can be), and it fails to satisfy (HN). So (HN) does not capture a metaphysically necessary condition of teleology.

The interlocutor might want to suggest that (HN) is not meant to be a metaphysically necessary condition, but some weaker kind. But I don't see what kind of weaker condition it could be. One cannot claim that (HN) is merely nomologically necessary, because (HN) is actually false – that's not how stones move in our world.[6] (HN) contradicts the actual laws of nature, so it describes a nomologically impossible phenomenon. Moreover, if one wants to capture the essence of teleology, mere nomological necessity does not fit the bill in the first place.

Counterexamples aside, there is a much more fundamental problem with the Hawthorne/Nolan conception: it is hard to generalize to final states that are not mere spatial positions. Louis strives to conquer Prussia – is his behaviour is teleological at *t* iff his action at *t* is part of the fastest possible campaign against Prussia? Not necessarily. Deliberate delay could, in principle, contribute to a more decisive victory. Is a foetus teleologically directed at becoming a healthy baby iff its biochemical processes turn it into a healthy baby as fast as possible? Again, no. So teleology is not locally optimal end-velocity.

5. Bedau's Beneficial Causes

The philosopher of biology Mark Bedau has offered a highly interesting theory according to which biological teleology involves the good of the organism. Although Bedau did not aim to define spooky teleology in general, his conception can be turned into such a definition by treating goodness as a metaphysically heavyweight general requirement.

Bedau's account is designed to cover functional traits like the heart's beating to circulate blood. His core idea is the following:

(B) *x F*s in order to *G* iff *x F*s because [*F*-ing contributes to *G*-ing & *G*-ing is good][7]

For example, the heart beats in order to circulate blood iff the heart beats because [its beating contributes to blood circulation & blood circulation is good for the organism]. Goodness is directly involved in teleological processes on this account; it helps explain why the process in question occurs.

(B) is not spooky by default, because its definiens is logically compatible with axiological antirealism. It could be that (B) is true but goodness is only in the

[6] For example, if you release a stone at the International Space Station, it will not satisfy (HN).

[7] Things enclosed in [...] are meant to be complex facts. "*G*-ing is good" is elliptical for "*G*-ing is good for the organism that has *x* as a part." Note that the definiens of (B) is very different from the following claim: (*x F*s because *F*-ing contributes to *G*-ing) & (*G*-ing is good). For discussion, see Bedau (1992. 787ff).

eye of the beholder. In that case, teleology itself is in the eye of the beholder. Alternatively, a reductive naturalist account of value turns (B) into a reductive naturalist analysis of teleology. (Of course, values must be naturalized without mentioning teleology to make this work.)

On the other hand, if (B) describes a real, observer-independent phenomenon *and* goodness cannot be reduced to value-free facts, then (B) is definitely ***spooky***. According to mainstream physicalism, irreducible objective values are not implicated in natural processes.

Whether (B) is ***idiomatic*** depends on the metaphysical content of "because". I can think of two options here: the "because" of causal explanation and the "because" of metaphysical explanation. The former makes (B) ***unidiomatic***, as it is hard to see how values could cause things. (Values can be treated as causes in the philosophy of action, but that's beside the point here unless one opts for mentalism.) Alternatively, "because" could signal some kind of 'metaphysical explanation' that is, according to Fine (2001) and others, symptomatic of metaphysical grounding, an allegedly all-important relation that plays a theoretical role similar to, but more general than, supervenience. This interpretation, however, renders the resulting analysis highly dubious. To illustrate, consider the claim that the the heart beats because$_{metaphysical}$ [its beating contributes to the circulation of blood & blood circulation is good]. If "because$_{metaphysical}$" signals metaphysical grounding, then the fact that [the heart's beating contributes to the circulation of blood & blood circulation is good] grounds the fact that the heart beats. But that sounds wrong – grounding is linked to phenomena like composition, realization, and constitution (cf. Schaffer 2009, 2017). None of those fit the relation between blood circulation and the heart's beating, which is a paradigmatically causal one.

So (B) is not ***idiomatic***. Nor is it ***adequate***, because robustly teleological processes can be directed at bad things. One can easily think of human history as a march toward a giant catastrophe, as a teleological process aimed at evil and destruction.[8] Yet under (B), this is logically impossible. Further, teleological processes whose ends are neither good nor bad are also impossible according to (B), even though they are possible. For no particular reason whatsoever, Steve sets out to count the blades of grass in his garden. Steve's activity is clearly goal-directed, but its end state is neither good nor bad.[9] So goal-directedness is not essentially connected to value. (That said, (B) could be entirely adequate in the context of biological functions. My point is that (B) is not a *general* definition of spooky teleology.)

[8] See e.g. *The Dialectic of Enlightenment* by Adorno and Horkheimer (1947/2002).

[9] An interlocutor remarks that Steve may actually need the number for some reason, and if he does not, then he is probably obsessive-compulsive and so not a real agent. Such cases are all conceivable, but so is a scenario where Steve is neither obsessive nor has any further use or need for the number of blades of grass in his garden. It is possible for people to strive for goals that have little or no value.

II. IRREDUCIBLE PERSISTENT PLASTICITY

The rest of the paper will offer a definition of spooky teleology that is both old and new. More precisely, its teleological component is old and the spookiness factor is new. I use an approach that was popular in the 1950s but fell out of fashion afterwards.[10] I will first argue that the conception in question – the idea that teleology is persistent plasticity – is both ***idiomatic*** and ***adequate***. Then I will mix in spookiness by adding a clause about irreducibility.

Imagine placing a rat in a maze that has food at the center. If the rat's sense of smell is good and the center of the maze can be reached from every entry point, then the rat will navigate to the center regardless of its initial position. Its goal-directed behaviour is *plastic* – moderate perturbations of the initial state result in virtually no perturbation of the final state.

The rat's behaviour is also *persistent* in the following sense: if we block a passage that the rat wants to traverse, then the rat will find an alternative route. If we place hurdles in its way, it will climb over them. And so on. The rat corrects its course when it encounters obstacles.

More generally, a behavioural disposition *D* is *plastic* with respect to state *G* iff bearers of *D* are disposed to reach *G* from a wide variety of starting points (so that moderate perturbations of the initial state result in virtually no perturbation of the final state). And *D* is persistent iff bearers of *D* are disposed to course-correct in the face of obstacles. *D* is persistently plastic iff it is both persistent and plastic.

Building on a venerable tradition in the philosophy of biology, I suggest identifying teleology in general with persistent plasticity:

(T) *x*'s goal is to reach state $G =_{df} x$ has a plastic and persistent disposition to reach *G*

This definition is ***idiomatic***. Apart from the concept of dispositions, which is familiar, (T) only uses the concepts of persistence and plasticity, which are easy to grasp and do not involve any unfamiliar metaphysical machinery.

Showing that (T) is ***adequate*** will take a bit more work. I submit that the definiens is a necessary condition of teleology. Surely, whatever a teleological system is, it ought to behave in a plastic and persistent manner with respect to its goals. If the rat is disposed to reach the food from only one entry point and it starts scratching its ear in all other cases, then it is not teleologically directed

[10] My main inspiration is Braithwaite (1947), complemented with Nagel's (1977. 272) notion of persistence. This is an amalgam of what Garson (2016. 17) calls the "behaviouristic" and the "mechanistic" approaches to teleology, the latter of which was influenced by cybernetics. Both were popular in the 1950s but are seldom defended today in the philosophy of biology (with the exception of McShea and his coworkers, see later).

at feeding. Similarly, if the rat stops when it encounters obstacles that it could overcome, then it is probably not hungry.

So the question is whether (T) states a sufficient condition of goal-directedness. According to Garson (2016. 23), (T) is threatened by overbreadth, because it classifies a wide variety of non-teleological systems as teleological. Imagine a marble ball that is released at the edge of a semi-spherical bowl. The ball ends up at the bottom of the bowl, and it does so in a plastic way: no matter where we release it, the ball always reaches the bottom. Its behaviour is persistent as well, because "it can adjust its trajectory in the face of obstacles" (Garson 2016. 23). So the ball has a goal, according to (T). But it does not. So (T) is wrong.

This objection seems to me quite easy to neutralize, because the ball's behaviour is obviously *not* persistent. If we place barriers on the inner surface of the bowl (so that it resembles a maze), the ball will get stuck on its way down. Unlike the rat, it is manifestly *not* disposed to course-correct.

One could modify the counterexample by reference to some brute physical disposition that is impossible to obstruct. Suppose (for the sake of argument) that the Sun's gravitational pull on Earth is the manifestation of a disposition. The disposition in question is plastic: wherever you put the Earth, the Sun will attract it. And the disposition in question is persistent: no matter how you try to obstruct it, the Sun will always attract the Earth. So, under (T), the Sun's goal is to attract the Earth. But the Sun has no goals. So (T) is wrong.

This counterexample also fails, because the concept of obstruction has no meaning with respect to gravity. There are no anti-gravitational shields. The claim that gravity course-corrects in the face of obstacles, if true, is merely vacuously true, like the claim that all unicorns love jazz. Accordingly, the counterexample can be neutralized by making explicit the presupposition that *x* has a *nonvacously* plastic and persistent disposition to reach *G*. Note, further, that if one considers a fundamental physical phenomenon that *can* be obstructed, like electromagnetism, then (T) will correctly classify it as non-teleological. Electric forces do not get around nonconducting materials that are placed between charged bodies.

Woodfield raises an objection about the individuation of goals in conceptions similar to (T). Specifically, he claims that any behavioural disposition can be rendered plastic by choosing a sufficiently fine-grained level of description:

> [A]ny straight line [in configuration space] representing a causal process can be bulged out in the middle or opened up like a fan, simply by choosing more refined criteria of type-identity for the middle or the start of the chain than for the end. (Woodfield 1976. 46)

Consider a defective rat that reaches the food from initial state I_1 only, aimlessly wandering around in hundreds of other possible cases. We can increase the number of goal-attaining initial positions if we switch to a finer-grained descrip-

tion: take the state I_1*, which differs from I_1 by a single molecule. The rat will also reach the food from I_1*. Now we have *two* favourable initial states, and we are on the road to plasticity. By choosing an appropriate level of grain, we'll be able to claim that the rat finds the food from a large number of initial states.

In response, one can point out that choosing different levels of grain to describe the end state and the trajectories leading up to it does not make the behavioural disposition in question plastic any more than calling tails "legs" makes dogs five-legged animals. Such differential tinkering is completely unmotivated from an epistemological standpoint, and it is easily circumvented by requiring that the final state and the activity leading up to it be described at the same level of grain (since they belong to the same level of reality).

Scheffler (1959) calls attention to 'the difficulty of multiple goals' in connection with definitions like (T). Suppose that some event E always accompanies the end state G – *s*ay, the rat always touches the food with its snout before feeding, or it always defecates after feeding. Then any behaviour that is plastic and persistent with respect to G will *ipso facto* be plastic and persistent with respect to E, and so (T) will classify E as a goal, even if it is a mere by-product.

This objection also fails. If E is some part of G (e.g. E is the state of the rat's touching the food with its snout), then E can plausibly be taken to be part of the goal. On the other hand, if E is distinct from G, then one can engineer a situation where E is absent but G is still attained (for example, one can prevent the rat from defecating after feeding).

To undermine this response, the interlocutor would have to find an E that is distinct from G yet necessarily accompanies G. To illustrate, let E be the rat's heart continuing to beat; surely this must accompany any instance of feeding. (T) then tells us that is part of the rat's goal to have its heart continue to beat while it eats. Or let E be the food's exerting a tiny gravitational pull on the rat; this, too, must accompany any instance of feeding, and so by (T), it is part of the rat's goal that its food exert a gravitational force on it.

But such counterexamples are also easy to dispose of. Whatever life-permitting situation the rat finds itself in, its heart will beat and external objects will exert gravitational forces on it. These general background conditions can and should be disregarded when we look at the behavioural tendencies that are present in a given situation. If the existence of such nomologically necessary background conditions were sufficient to undermine (T), then theories of causation and the ontology of dispositions would be similarly undermined, since the individuation of events is equally important there. Breaking is always accompanied by tiny gravitational interactions between the broken fragments; does it follow that a wine glass, apart from being disposed to break, is also disposed to have its fragments exert a minute gravitational pull on each other? And does this create an insurmountable problem for the individuation of fragility? I don't think so. The precise solution to this problem is far from obvious, but the problem is so

general that it cannot constitute an argument against (T) unless the interlocutor is willing to claim that the metaphysics of nature is in general undermined because of this single issue.

I conclude that (T) is ***adequate***. It is, however, not yet ***spooky***. I propose to make it so by adding a clause about irreducibility:

(ST) *x* is characterized by spooky teleology with respect to goal $G =_{df}$ (i) *x* has a plastic and persistent disposition *D* to reach *G*, and (ii) *D* is not reducible to the properties of *x*'s parts or to properties of things distinct from *x*

In a slogan: spooky teleology is irreducible plastic persistence.

This definition is still ***idiomatic***, since the idea of reducibility (understood as an ontological, not linguistic or epistemological, phenomenon) is quite familiar. (ST) is also ***adequate***, since it defines teleology the same way as (T) did: clause (i) is simply the definiens of (T). Finally, the definition is distinctly ***spooky***, because standard contemporary physicalists are likely to reject the idea that plastic and persistent behavioural dispositions (such as a rat's tendency to navigate to the food) are irreducible. The only irreducible dispositions are the fundamental physical ones, according to standard contemporary physicalism; everything else is just a jumble of fundamental interactions.

A reviewer complains that (ST) is not too different from 'minimum energy' conceptions like the Hawthorne–Nolan one. And similarities aside, why should we regard the processes in question as goal-directed at all? As the reviewer remarks, "by these lights, even a waterfall would be goal-directed, even if the pool at the bottom is certainty not 'directing' the water into it."

As to the first worry, (ST) is not a minimum-energy principle simply because it does not mention energy at all. According to (ST), an entity or system can be goal-directed even if it chooses paths that do not satisfy a 'least-action' law. A military commander who wins a battle but loses more soldiers than the best of all commanders would have is nonetheless behaving teleologically if he satisfies the definiens of (ST).

As to the second worry, I reply that waterfalls are not goal-directed under (ST), because they are neither persistent nor plastic. Moderate perturbations of their initial conditions do not result in virtually no perturbations in the outcome – if the river is diverted at the origin, it will not curve back to form a waterfall at the same cliff. Nor is a waterfall persistent. If a dam is erected right at the cliff, the water will not flow around it to reach the pond.

Apart from being ***idiomatic***, ***adequate***, and ***spooky***, (ST) has two further advantages. First, it fits the intuition that in the case of robustly teleological processes, the goal is the metaphysical ground of the activity that occurs for its sake. To see why (ST) supports this intuition, consider the most extreme form

of spooky teleology conceivable under (ST), the case where *x*'s plastic and persistent disposition is not reducible to *anything* – not even to other properties of *x*. Call this "maximally spooky teleology." And consider the variety of fine-grained dispositions that *D* entails in such cases: *x* will reach G under initial condition I_1, *x* will reach *G* under a rather different initial condition I_2, *x* will reach *G* if faced with obstacle O_1, *x* will reach *G* faced with a different obstacle O_2..., where I_1, I_2... and O_1, O_2... are gerrymandered collections, not instances of some natural kind. The only thing that unifies this motley bunch of behavioural tendencies is *G*, the goal. And so if *D* is not reducible to anything, then the only metaphysical structure that is discernible in *D* is that its manifestations all terminate in *G;* the goal *G* individuates *D*. So if teleology is maximally spooky, then the the metaphysical explanation of the activities leading to *G* will always involve *G* itself, and in that sense, the goal will ground the activities that occur for its sake.

Another valuable feature of (ST) is that it treats spooky teleology as a subtype of teleology, since (ST) includes the definiens of (T). (T) itself is compatible with reductionism, so (ST) makes it possible for friends of spooky teleology to conduct empirically informed debates with friends of naturalized teleology. Moreover, this is not just a logical possibility but an existing opportunity, since the philosopher of biology Daniel McShae and his coworkers do use (T) in their reductive analyses of biological teleology (McShea 2012, Lee–McShea 2020, Babcock–McShea 2021). Friends of spooky teleology are off to a good start if they manage to show that those reductive projects fail.

More generally, friends of spooky teleology who accept (ST) can deploy preexisting arguments for emergence in the philosophy of science to resurrect real teleology. If organisms, or ecosystems, or human minds, or plural subjects, or societies etc. are (i) strongly emergent and (ii) display persistently plastic behaviour, then spooky teleology is a reality, according to the definition that I recommend. And since arguments for (i) and (ii) are not hard to come by, (ST) supports the interim conclusion that robust teleology may not be quite as dead as it seems.

III. SUMMARY

This paper sought a definition of the kind of teleology that modern science allegedly eradicated. My inquiry was ahistorical: instead if trying to reconstruct old doctrines, I was looking for a definition suitable for contemporary analytic metaphysics; one that is ***idiomatic*** (intelligible for contemporary metaphysicians), ***adequate*** (defines goal-directedness), and ***spooky*** (describes a phenomenon that standard contemporary physicalists do not believe in).

I introduced six candidate definitions, and I argued that their scorecard is the following:

	Idiomatic	*Adequate*	*Spooky*
Mentalism	+	–	+
Retrocausality	+	–	+
Causal Powers	+	–	–
The Hawthorne/Nolan Model	+	–	+
Bedau's Beneficial Causes	–	–	+
Irreducible Persistent Plasticity	+	+	+

The last conception is superior to all the others, in my view, and it justifies the intuition that in cases of maximally spooky teleology, the goal is one of the metaphysical grounds of the activities leading to it. I also indicated that my proposed definition helps stage empirically informed debates with friends of naturalized teleology. The outcome of those debates – the viability of resurrecting robust teleology – will depend on the strength of emergentist arguments in biology, the philosophy of mind, and social ontology.

REFERENCES

Adorno, Theodor – Max Horkheimer. 1947/2002. *Dialectic of Enlightenment*. Stanford, Stanford University Press.

Babcock, Gunnar – Daniel W. McShea. 2021. An Externalist Teleology. *Synthese*. 199. 8755–8780. https://doi.org/10.1007/s11229-021-03181-w

Bedau, Mark. 1992. Where's the Good in Teleology? *Philosophy and Phenomenological Research*. 52/4. 781–806. https://doi.org/10.2307/2107911

Braithwaite, Richard B. 1947. Teleological Explanation. *Proceedings of the Aristotelian Society* 47. i–xx. https://doi.org/10.1093/aristotelian/47.1.i

Corry, Richard. 2019. *Power and Influence: The Metaphysics of Reductive Explanation*. Oxford, Oxford University Press. https://doi.org/10.1093/oso/9780198840718.001.0001

Descartes, René. 1641/2008. *Meditations on First Philosophy*. Oxford, Oxford University Press.

Dupré, John. 2012. Review of Thomas Nagel's *Mind and Cosmos*. *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/mind-and-cosmos-why-the-materialist-neo-darwinian-conception-of-nature-is-almost-certainly-false/

Einstein, Albert – Max Born – Hedwig Born. 1971. *The Born-Einstein Letters*. London, Macmillan.

Fine, Kit, 2001. The Question of Realism. *Philosophers' Imprint*. 1/2. 1–30.

Garson, Justin. 2016. *A Critical Overview of Biological Functions*. Berlin, Springer. https://doi.org/10.1007/978-3-319-32020-5

Hawthorne, John – Daniel Nolan. 2006. What Would Teleological Causation Be? In J. Hawthorne. *Metaphysical Essays*. Oxford, Clarendon Press. 265–284. https://doi.org/10.1093/acprof:oso/9780199291236.003.0015

Hobbes, Thomas. 1651/1998. *Leviathan*. Oxford, Oxford University Press.

Kelemen, Deborah – Joshua Rottman – Rebecca Seston. 2013. Professional Physical Scientists Display Tenacious Teleological Tendencies: Purpose-Based Reasoning as a Cognitive Default. *Journal of Experimental Psychology: General*. 142/4. 1074–1083. https://doi.org/10.1037/a0030399

Koons, Robert C. 2021. The Ontological and Epistemological Superiority of Hylomorphism. *Synthese*. 198. S885–S903. https://doi.org/10.1007/s11229-016-1295-6

Koons, Robert C. – George Bealer (Eed.) 2010. *The Waning of Materialism*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199556182.001.0001

Lee, Jong Gwan – Daniel W. McShea. 2020. Operationalizing Goal Directedness: An Empirical Route to Advancing a Philosophical Discussion. *Philosophy Theory and Practice in Biology*. 12/5. https://doi.org/10.3998/ptpbio.16039257.0012.005

McKitrick, Jennifer. 2018. *Dispositional Pluralism*. Oxford, Oxford University Press. https://doi.org/10.1093/oso/9780198717805.001.0001

McShea, Daniel W. 2012. Upper-Directed Systems: A New Approach to Teleology in Biology. *Biology and Philosophy*. 27. 663–684. https://doi.org/10.1007/s10539-012-9326-2

Molnar, George. 2003. *Powers*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199204175.001.0001

Nagel, Ernest. 1977. Teleology Revisited: Goal Directed Processes in Biology and Functional Explanation in Biology. *Journal of Philosophy*. 74. 261–301. https://doi.org/10.2307/2025745

Ott, Walter. 2009. *Causation and Laws of Nature in Early Modern Philosophy*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570430.001.0001

Paolini Paoletti, Michele. 2021. Teleological Powers. *Analytic Philosophy*. 62. 336–358. https://doi.org/10.1111/phib.12245

Rose, David – Jonathan Schaffer. 2017. Folk Mereology is Teleological. *Noûs*. 51/2. 238–270. https://doi.org/10.1111/nous.12123

Schaffer, Jonathan. 2009. On What Grounds What. In D. Chalmers – D. Manley – R. Wasserman (eds.) *Metametaphysics*. Oxford, Oxford University Press. 347–83. https://doi.org/10.1093/oso/9780199546046.003.0012

Schaffer, Jonathan. 2017. The Ground Between the Gaps. *Philosophers' Imprint*. 17/11.

Scheffler, Israel. 1959. Thoughts on Teleology. *British Journal for the Philosophy of Science*. 9. 265–284. https://doi.org/10.1093/bjps/ix.36.265

Trommler, Friederike – Marcus Hammann. 2020. The Relationship between Biological Function and Teleology: Implications for Biology Education. *Evolution: Education and Outreach*. 13/11. https://doi.org/10.1186/s12052-020-00122-y

Woodfield, Andrew. 1976. *Teleology*. Cambridge, Cambridge University Press.

Mohsen Moghri

An Axiological Ultimate Explanation for Existence*

Why is there anything concrete at all, instead of there being nothing concrete? Leibniz (1714/1989) devised this puzzle (hereafter "the Why question") as a way to uncover what constitutes the nature of our concrete world.[1] Leibniz's own suggestion, after elaborating on the Why question, was that our concrete world is the best possible world and its Goodness is the reason it exists. Following Leibniz, Kuhn (2007) illuminates how an explanation for all concrete existence helps one discover the fundamental feature of our universe, and Holt (2012) adds that only by asking why there is a concrete world can one know why the universe behaves in a certain way. Further, Nozick (1981) intensifies the latter point by maintaining that without an answer to the Why question, one might not be able to answer any other question at all.

One might attempt to solve Leibniz's puzzle using one's suggested answer to similar fundamental questions. But many such responses fail to solve the *ultimate* Why question. Consider for example the question of why there is this universe – a universe devoid of seemingly simple laws that lead to the existence of living beings like us – rather than another universe. The benevolent omnipotent God of Theism or the fundamental Laws of Nature might answer the question of why our universe is fine-tuned for life, but the question still remains as to why the fine-tuner of our universe itself or himself exists. In this respect, Leibniz emphasized that a proper answer to the Why question must have a stopping point; it must be ultimate. He sought "*a sufficient reason that has no need of any further reason* – a 'Because' that doesn't throw up a further 'Why?'…" (Leibniz 1714/1989). Note that while an explanation for the existence of all concrete things might be complete by leaving no concrete thing unexplained, it might fail to be ultimate: The question may still remain unanswered of why the explana-

* Thanks to the anonymous referees of this journal, the audience at a February 2021 online seminar at the Center for Religious Studies at the CEU, and helpful comments from Daniel Kodaj on the earlier drafts of this article. This work was supported by the UKRI-Horizon Europe Guarantee [grant reference: EP/X022633/1].

[1] I take concreteness, here, in terms of capability to involve in an efficient causal relation.

tion for the existence of every concrete thing or for the whole of them obtains. When an explanation for something is both complete and ultimate, everything is explained concerning that thing and no brute fact remains. Swinburne (2004) calls such an explanation an "Absolute Explanation" and suggests that there cannot be one such explanation for the existence of our world. Swinburne argues that God created the whole world. However, he accepts God not as the ultimate explanation, but as "the ultimate brute fact."

But isn't there a way to escape all existence from being ultimately brute? The existence of our world seems far away from being logically necessary; neither does it seem to be ultimately explainable using logically necessary truths (cf. Rowe 1970, van Inwagen 1983). An alternative is, nevertheless, to explain concrete existence through beings or facts that are metaphysically necessary; they bear their explanation within the essences involved. Accordingly, many scholars suggest that the existence of God or obtaining of the Laws of Nature are metaphysically necessary and ultimately explain all concrete existence (cf. Hawking & Mlodinow 2010, Loewer 2012, Lange 2014, O'Connor, 2008, Pruss, 2006).[2] In disagreement, however, some contend that metaphysically necessary facts, even some logical or mathematical truths, may further be explained (Van Cleve 2018, Vintidals 2018). It seems that we are still entitled to ask why some metaphysically necessary facts obtain and not others. If metaphysical necessities are apt for further explanation, many ambitions, in theology and philosophy of science, to reach an ultimate explanation for the existing world are not fulfilled.

As an explanation that could block the chain of further explanations, one might suggest that some fact literally explains itself by being an instance of itself. For example, Nozick (1981) argues that the Principle of Fecundity, the principle that states "All possibilities obtain," is an instance of itself; the principle of Fecundity subsumes itself because the obtaining of that principle itself is a possibility. While the latter kind of self-explanation is generally regarded as dubious, I develop an alternative self-explanatory account that may succeed in providing an ultimate explanation for existence. Section 1 shows that a viable explanation for all concrete existence should proceed not causally but teleologically, and appealing to abstract facts of value provides a suitable candidate. In Section 2, I elaborate on the conditions that require an ultimate explanation and on what might constitute such an explanation. In section 3, a variant of self-explanation, namely self-subsumption – obtaining of a fact literally in virtue of itself – is introduced with an exploration of the main objections to it. Finally, in chapter 4, I construct a self-subsuming ultimate explanation that avoids the latter objections.

[2] Although other terms are used for the necessity of Laws of Nature, such as Nomological or Natural necessity, this kind of necessity can be regarded as a metaphysical necessity. For, Laws, on this view, are certain generalizations among concrete entities whose necessity follows from the natures of those entities (Bird 2007).

I. EXPLAINING CONCRETE EXISTENCE BY ABSTRACT FACTS OF VALUE

There seems to be no logical contradiction in supposing that concrete things in the world vanish altogether. In other words, it seems that the non-existence of the whole world is logically possible. For one thing, even the advocates of the Ontological Argument have retreated from insisting that a *logically* necessary being exists. Besides, arguments for the logical necessity of existence have not been very convincing. Therefore, one may sensibly follow Leibniz in asking: Why is there something concrete, instead of there being nothing at all?

Despite this, many scholars have argued that Leibniz's question is meaningless or without an answer. Most of them object that every concrete thing has a causal explanation inside the world, and there cannot possibly be something that causes the whole world to exist. Every possible explanation for the existence of concrete things, they say, would be part of the whole world; that explanation thus needs another explanation that would itself be part of the whole world again, *ad infinitum*. Following Hume, most critics conclude that the existence of our world does not need any explanation whatsoever.

But isn't it mysterious that the world behaves in such an orderly way? Further, why isn't there a different order of concrete things in the world? More importantly, there might have been no order, or worse, nothing instead. If it is rational to ask why some or other concrete thing exists, why not demand an explanation for the whole of existence? To be sure, it is logically possible that the world's existence is a brute fact, having no explanation. However, as long as no evidence is provided that the world's existence has no explanation, one should suppose there is an explanation. For, it is widely argued among philosophers that everything must have an explanation unless some evidence shows why there cannot be an explanation.

Yet, it is a mistake to restrict all explanations to causal effects. This restriction stems from the traditional idea that a cause must be homogeneous with its effect. However, just as matter can turn into energy and living organisms can originate from inorganic components, so too might concrete realities originate from non-concrete ones. We seem to have not enough reason in claiming: All explanations of concrete things need earlier concrete things as their causes. The source of the whole existing world might fall outside the category of concrete existence. An abstract fact might explain why our world exists.

The latter kind of explanation can be interpreted teleologically, as well. If our world has a special feature, such as extreme simplicity or abundance, it is rational to believe that the world is actual because it has such a unique feature. It is extremely unlikely that a special feature in our world arises as a matter of happenstance. Yet, that feature does not need to be instantiated in a previously existing thing; neither does it need a preceding agent to make it actual. Instead, as Platonists maintain, having some abstract feature (a Form) might have led

something possible into being actual. Leibniz asserts, in this respect: "[F]rom the very fact that there exists something rather than nothing, it follows that in possible things, or in possibility or essence itself, there is a certain need of existence..." (Leibniz 1898. 340). Here, the abstract feature in possibility or essence is not the efficient cause, but the final cause of existence: Because of having some special features, certain possibilities seek to be actual.

It may still seem very odd that abstract facts explain concrete existence, but that kind of explanation is in fact what many theists and scientists use as well, most of them without even knowing it. Stephen Hawking (2010), for example, takes the highest degree of symmetry involved in the fundamental laws of quantum as an indication that those laws must have governed the world. Here, the reason why our world exists is supposed to be the abstract fact that its fundamental laws are super-symmetric. As an example among theists, William Lane Craig (2008) argues that God's explanation resides in his own nature. He considers God as self-caused but not in the sense that he is the efficient cause of himself. Instead, the essence of God is, according to him, so perfect that God requires no efficient cause to exist. That is God's essence that requires him to exist. But is this non-efficient cause, this perfection in God's essence, not an abstract factor that explains why God exists? Here, as well as in many forms of the Ontological Argument, what really explains God's concrete existence is an abstract fact: That the essence of God is maximally great, good, or valuable. Both Hawking and Craig use the same force that they exclude from abstract things in order to explain concrete existence.

The term, Axiological explanation, is what John Leslie (1979) and Nicholas Rescher (1984) use when the value of something explains its existence. They both insist that the Axiological explanation does not require the mediatory act of an agent, such as God. If value alone can be said to explain why God exists, why not use value instead to explain the whole concrete world without entering God? Leslie and Rescher follow Plato's footsteps in viewing the Form of the Good as what gives existence to the world. Leslie goes so far as to use the term, God, for the creative force of value. Some contemporary theists also explicitly express a similar approach. Paul Tillich (1951), for example, maintains: God "is not a being" but is instead "the power of being", "the creative ground of existence." While there is some power inside all valuable possible worlds to exist, a world with the most value exceeds others in its power and becomes actual.

The teleology involved here may well be termed naturalistic: a goal-directedness without an agent's purpose. While a purpose must be somebody's purpose, a value can be totally impersonal. Here is some evidence: First, the real, inherent value of something can be underestimated or overestimated. Second, just as a lawful world ripe for the existence of free intelligent beings is really better than there being nothing, nothingness would be really better than a world full of chaos and agony. In the case of nothingness, there would be no agent to

determine a value according to some purpose. So, the value of a world may, in itself and without being the value for someone or something, be the sufficient reason for that world's existence. Rescher describes this naturalistic teleology as follows: "Reality is inherently disvalue-phobic – or value-tropic if you prefer. But this transpires only as a matter of a strictly natural process: preference and purpose have nothing to do with it. The "aversion" of reality to disvalue is as natural as the "aversion" of one magnetic pole to another of the same polarity" (Rescher 2010. 72).[3]

It is worth mentioning here that there is no agreement on the features that make a world intrinsically valuable. Leslie's focal point is on Ethical value, though he uses that term in a much wider sense than just the moral actions of agents; it includes even the world's lawfulness. Rescher, in contrast, focuses on Cosmic values of order, simplicity, and variety. Nevertheless, he argues that the combination of Cosmic values is optimized only if there is an evolutionary process leading to the emergence and welfare of intelligent beings. In spite of these differences, neither Leslie nor Rescher appeals to an agent's purpose to explain why valuable things are actual.

Even so, the Axiological explanation has been subject to many criticisms. If value rules over the world, shouldn't we expect that the world would be full of prosperity and happiness? Why then is instead evil everywhere? One might appeal here to a theistic response to the Problem of Evil: Valuable things are often in contrast with each other; it is impossible to have all good things in the world. In that case, we should at least expect that our world would be the most valuable one, the best of all possible worlds. On the contrary, however, the claim that our world is the best possible world has not convinced many philosophers (Mackie 1982, Parfit 1998), not even some theists (Plantinga 2011). Besides, there are some difficulties in determining the best possible world. As mentioned above, the overall value of a world is achieved through a combination of several features, such as simplicity and variety. In that scenario, more than one combination may lead to the most valuable outcome. Some difficulties as well arise about supposing that there is one highest level of value for a world. The overall value of a world may simply be multiplied.

Moreover, what evidence do we have that value governs the world? The world's lawfulness, having a good combination of simplicity and complexity, and

[3] An anonymous referee of this article argues that attributing a creative force to an ontological realm of value, devoid of any intention or conscious purpose, is still highly implausible. However, the idea has a strong historical precedence, especially in Platonic and Neoplatonic traditions. Additionally, there has been a recent revival of Aristotelian Metaphysics in analytic philosophy to explore the nature and principles of non-causal explanations, using constitutive parts or features of existing things, among which their values might be included. The latter kind of explanation is the subject of grounding theories, and the author is currently working on developing the Axiological explanation in conjunction with those theories.

fine-tuning for intelligent life may provide some evidence that value rules over our world. However, these are claimed to provide the same amount of evidence for other hypotheses as well. For example, a Multiverse might exist in which each world is governed by a different natural law. This makes the existence of every universe the least arbitrary (Unger 1984). In the latter hypothesis, it is not very surprising that one among those universes has the fundamental law of our universe and leads to intelligent life.

Despite these difficulties, one feature of the Axiological explanation may provide a decisive advantage over other alternative explanations. Both Leslie and Rescher argue that only value can provide an ultimate explanation for existence, an explanation that does not provoke a further why question. If one invokes God's creative act to explain why the world exists, another may further explain why God exists. Similarly, the fundamental laws of the universe (or the Multiverse) may further be explained. On the other hand, facts of value are self-explanatory, in the sense that it lies in the nature of good things that they are valuable. To ask why something has value is like asking why redness is nearer to being purple than to being blue. Not even Omnipotence could "give" intrinsic worth to anything, so Leslie (2014) points out. Facts of value are synthetically, metaphysically necessary. Therefore, no further explanation is needed, according to the advocates of the Axiological explanation, for why value obtains. They claim that, in explaining existence, only facts of value can escape the explanatory regress. Thus, the higher explanatory power of the Axiological explanation indicates, for them, that it is more likely to be true.

Nevertheless, here comes a bigger problem: Many scholars argue that God or the Laws of Nature, too, are metaphysically necessary. They claim that it lies in the nature of concrete objects that they behave lawfully, or in the nature of God that He exists. Considering that, there seems to be no privilege in explaining existence through facts of value. For, no extra amount of explanatory power has been provided through the Axiological explanation. Even more, we saw in the case of a metaphysically necessary God that it makes sense to ask: Why is God metaphysically necessary? Why does the essence of God explain His existence? In effect, there are attempts to place the necessity of God in facts of value. Similarly, some advocates of the metaphysical necessity of natural laws demand an explanation for why those necessities obtain (cf. Lowever 2012) If the metaphysical necessity of God or the Laws of Nature needs to be explained by some or other fact, one might still wonder why that which has maximal value necessarily exists. Some attempts to place the necessity of facts of value in God further confirms the latter point (cf. Craig 2008, Murphy 2011).

In contrast to all of these attempts, many deny that metaphysical necessities could have further explanations. John Heil (2018), for example, asserts: There might be logically possible things that are not really possible; it is the nature of reality that determines what is really possible. Metaphysical necessities, ac-

cording to him, could not be otherwise because other alternatives are not really possible; therefore, they are not apt for an explanation, and neither could one coherently consider them as brute facts.

For sure, some fact's being metaphysically necessary means that no alternative fact is really possible. Yet, it does not entail that no other fact forces the first fact to obtain. Physical objects have no alternative other than to attract each other, but there are still some facts, the fundamental Laws of Nature, that explain why objects attract each other. As another example, some theists argue, created things exist with a metaphysical necessity. Still, it is a metaphysically necessary creator that explains why created things necessarily exist. In an explanation, in fact, we are not trying to show only why something is not otherwise, but also why it is as it is (cf. Murphy 2011) Therefore, metaphysical necessities, such as facts of value, might have obtained by virtue of an explanation.[4]

II. NECESSITY AS AN ULTIMATE EXPLANATION

To elucidate what counts as an ultimate explanation for all concrete existence, Leibniz further maintained that the only viable answer to the Why question is one that involves something self-explanatory "carrying the reason for its existence within itself" (Leibniz 1714/1989). He suggested a necessary being for that purpose. However, it is generally accepted now, even by some proponents of the Ontological Argument (for example, Plantinga 1977), that it is not clear at all that a necessary being is really possible. Following Hume, many contemporary thinkers accept that it is *logically* possible that nothing existed. Moreover, some philosophers argue that explaining the existence of our world through a logically necessary being results in necessitarianism, which contradicts the conviction that the world is contingent (Rowe 1970, van Inwagen 1983). As a new challenge to ultimately explaining all concrete existence through logical or mathematical necessities, Van Cleve (2018) and Vintidals (2018) argue that the obtaining of necessities may further be explained. Their main purpose is to show that it is possible to accept some necessities as brute facts and leave them unexplained. By arguing that necessities may be apt for further explanation, their argument,

[4] Not only metaphysical necessities but also some mathematical necessities are viewed as apt for further explanation. There is a distinction, among Mathematicians, between proofs that only prove and proofs that explain (cf. Mancosu 2001), and Lange (2014) argues that such a distinction is not a mere subjective distinction; it denotes something real. When mathematicians use Peano axioms to prove 2+2=4, they consider those axioms as some reality in the world that explains why such a necessary fact obtains. To see that the relationship here is not a mere entailment, but an explanation, one should note that the reverse does not work: 2+2=4 cannot prove Peano axioms. A necessary fact is therefore explained by another necessary fact.

at the same time, cast further doubt on there being an ultimate explanation for all concrete existence in the domain of logical necessities.[5]

What about explaining concrete existence through metaphysical necessities? Can those alternative necessities provide ultimate explanations? Fine (2012) defines metaphysical necessity as the obtaining of something in virtue of its own essence or nature. One can accordingly interpret metaphysical necessity in terms of ontological independence: A metaphysically necessary being, for example, would rely for its existence not on any other being, but on its own nature. Furthermore, Loewer (2012) defines metaphysical explanation of something in terms of its being grounded by more fundamental facts, whereas scientific explanations are generally considered in terms of prior events or facts. In these senses, if one explains the behaviour of concrete objects by natural laws but maintains that those laws are explained by virtue of the objects' essence, one must consider natural laws as metaphysically necessary.[6] God is also considered traditionally as a metaphysically necessary being. Craig (2008) explicates that when theists speak about a self-caused God, they mean not that God would exist prior to himself to cause himself, but that his essence (being perfectly good) is why he exists. Among these accounts of metaphysical necessity, one may regard a being, a law, or a fact as that which could block the chain of explanation by explaining itself. This sort of self-explanation must, so Van Cleve (2018) argues, be considered as an intrinsic, non-relational explanation, not as the case that some fact would literally explain its own obtaining.

Nonetheless, one should note, first of all, that the existence of things may primarily or ultimately be a brute fact. As Parfit (1998) contends, it is not impossible at all that our world exists as a matter of happenstance and, even if our world has an explanation for its existence, that explanation might have obtained without any further explanation. A world must, as a matter of logic, obtain somehow with or without concrete existing things, and it is possible that it contains concrete existence without any explanation. While many followers of Russell (1948) consider the existence of our world to be a brute fact, Swinburne (2004) and Carroll (2018) posit that the ultimate fact that explains all concrete existence is inexplicable. Meanwhile, Swinburne accepts God as the ultimate brute fact that is himself the explanation for all concrete existence. Similarly, Carroll argues that the highest-level scientific law or principle that explains other lower-level natural laws is without any further explanation, though he accepts that all concrete existence can be explained by natural laws.

[5] There is a distinction between epistemic and ontological brute facts (Barnes 1994). We are concerned here with the metaphysical and ontological aspects of explanation, not merely the ability of our cognitive faculties to *find* or to *know* an explanation.

[6] Lange (2014) calls natural laws "physically", "naturally", or "nomologically" necessary. He uses the term "metaphysical necessity" differently in terms of non-contingency and considers it in a group with logical and mathematical necessity.

Second, contrary to what Heil (2018) argues, one might consider even the obtaining of metaphysical necessities as subject to further explanation. Heil maintains that there might be logically possible things that are not really possible. According to him, the nature of reality determines what is really possible. Heil denies therefore that metaphysical necessities need further explanation. Since he regards metaphysical necessities as not apt for an explanation, he also denies that one could coherently consider metaphysical necessities as brute facts.[7] In contrast, I attempt to defend Rescher (2013) in thinking that Nature might allow for *many* real possibilities. Craig (2008) and Murphy (2011) argue that necessary facts and truths, especially moral necessities, may further be explained. Furthermore, in an extensive debate over an argument for the possibility of a world devoid of concrete objects,[8] many argue that an empty world was *really* possible (Efird & Stoneham 2005, Rodriquez-Pereyra 1997). If the latter arguments are successful, the possibility of an empty world is an instance of real possibilities that certainly have not been realized. So, there might be real possibilities that are not metaphysically necessary.

Rescher (2013) then argues that one must, as a methodological rule, prefer a real possibility that has an explanation rather than a brute real possibility. He maintains, in this respect, that to regard some fact as brute and inexplicable is our last resort. Many contemporary philosophers adhere to a weak version of the Principle of Sufficient Reason, which maintains that one must seek an explanation unless one finds a reason that there cannot be an explanation (Della Rocca 2010, O'Connor 2008). As a result, while there is no requirement for an explanation of all existence to be ultimate (nor is it required that it would have any explanation at all), to find an ultimate explanation is, as O'Connor (2008) argues, only a matter of adhering to a hypothesis with more explanatory power.[9] It is generally accepted, as a methodological rule in analytic philosophy, that a hypothesis involving more theoretical virtues (explanatory power and scope as well as simplicity among them) is more likely to be true.

Nor is it needed that an explanation of the existing world be contrastive at all. While some fact may explain the obtaining of A, that fact may not be able to explain why A obtains rather than B. For instance, the reason I chose tea to drink may be insufficient to explain why I chose tea over coffee to drink. Pruss (2006)

[7] Some philosophers maintain that metaphysical necessities are explained in terms of their own necessity. Rosen (2010), for example, claims that "whenever it is essential to x that p, p holds *because* it is essential to x that p." However, Van Cleve (2018) argues that this kind of essentialist explanation results in implausible explanatory regress because, in that case, every metaphysical necessity must be explained through a higher-level metaphysical necessity: For example, it holds that it is essential to x that p because it is essential that it is essential that p. Essentialist explanation is therefore, at least, as implausible as explanatory regress.

[8] The Subtraction Argument was firstly proposed by Baldwin (1996).

[9] Similarly, Swinburne (1997) accepts the hypothesis that God created the whole concrete existence in virtue of its utter simplicity.

and O'Connor (2008) argue that God, as a metaphysically necessary being, is the only ultimate explanation for the existence of all concrete things; though they concede that such an explanation cannot, and need not, explain why there is not nothingness instead. Thus, Goldschmidt (2011) speaks of a new Cosmological Argument for the existence of God based on the ability of the latter hypothesis to provide an ultimate explanation for all concrete existence. However, it is again a matter of explanatory power that one prefers an ultimate *contrastive* explanation of the existing world to a non-contrastive one. Therefore, if one finds an ultimate explanation that can explain why there is something concrete rather than nothing, it is rational to prefer the latter to all non-contrastive explanations of the existing world in virtue of its more theoretical virtues. Moreover, the theistic explanation of Pruss and O'Connor might not be enumerated among the ultimate explanations at all. For, as indicated before, even the metaphysical necessity of God may further be explained. In that case, the question is still unanswered as to why, among all real possibilities, it is the essence of God that makes him actual. Why did not another real possibility, for example, an empty world, obtained? Although the latter God-hypothesis may seem a complete explanation for the existence of all concrete things, it cannot therefore be ultimate in itself.[10] It seems that the prevalent theistic explanations must follow Swinburne's suggestion in accepting the existence of God as an ultimate brute fact.

On the other hand, as a kind of scientific self-explanation, accounts of a self-contained universe are what many have claimed to be the source of an ultimate explanation. The most prevalent account is the beginningless universe of Hume and some contemporary physicists. To strengthen this point, Quentin Smith (1988) argues that an infinite causal regress provides an account of a self-caused universe: Every concrete thing has its own cause within the universe. Alternatively, some theoretical physicists suggest a universe that literally causes itself in a circular process (Gott & Li 1998). One finds more recent suggestions for a self-contained universe in hypotheses such as the quantum gravity of Hawking (2010), quantum tunnelling of Vilenkin (2007), and quantum fluctuations of Krauss (2012), many of which result in considering a kind of Multiverse hypothesis (cf. Greene 2011).

How powerful are the latter suggestions in explaining existence? While most of those suggestions do not provide complete explanations for the existing world and leave the totality of all concrete existence (the existence of a whole Multiverse) unexplained, the beginningless universe involving infinite causes might be considered as explanatorily complete. For, every single event in that regress has its own explanation. For the same reason, Vintiadis (2018) argues that a regress of infinite explanations does not lead to a brute fact. Nevertheless, one

[10] One may still add some explanatory feature to a God-hypothesis in order to turn it into an ultimate explanation.

might argue that a causal or an explanatory regress result in unacceptable *circular* explanations (cf. Pruss 2006). A further problem reveals itself in the way of scientific ultimate explanations. Heller (2009) rejects various Multiverse hypotheses because he contends that they cannot be falsifiable. He argues then that one needs to add philosophical or theological explanations to provide an ultimate explanation for all concrete existence. Although I align myself with Greene (2011) in thinking that some of the previous scientific accounts are falsifiable, I attempt to argue that those scientific suggestions cannot solve Leibniz's puzzle. For, there still remains the question of why the supposed explanation of the existing world obtains. If one further explains why there is a self-contained universe, more explanatory power is provided than when a hypothesis considers the existence of a whole universe or Multiverse without any explanation.

To turn a Multiverse hypothesis into a complete explanation for concrete existence, natural laws must be regarded as metaphysical facts that require concrete things to exist (cf. Lange 2013). There are many critics, for example, Armstrong (1983) and Maudlin (2007), who argue that if natural laws have no metaphysical footing and are nothing more than Humean regularities in the behaviour of concrete objects, those facts cannot explain anything. Therefore, natural laws must be considered as *abstract* entities that non-causally explain why there is something concrete (cf. Brenner 2020, Moghri 2021). If one considers natural laws as concrete things that causally force other concrete things to exist or behave in a certain way,[11] the explanation of all concrete existence by natural laws cannot be complete until the existence of those laws is further explained.

Even though an explanation of all concrete existence using abstract natural laws might be complete, such an explanation does not still seem to be ultimate. Why do those fundamental natural laws obtain? In response, Loewer (2012) argues that the obtaining of natural laws can be grounded in what constitutes the essence of those laws – the very nature of concrete objects in behaving law-like. However, this seems to be a circular explanation to postulate that natural laws explain why there are concrete objects while the existence of those objects grounds the obtaining of natural laws. To remedy the apparent circularity in explanation, Loewer contends that explaining natural laws by concrete objects is a kind of metaphysical explanation while the reverse is a scientific explanation – explaining the existence of concrete objects by the force of prior laws. Thus, he attempts to escape the objection of circularity.

Nevertheless, Loewer's suggestion is still considered by Lange (2018) as an unacceptable self-explanation because of the transitivity between the two kinds of explanation. Lange might be correct in thinking that metaphysical and scientific explanations are linked by a transitivity principle. However, Loewer's argument

[11] Kuhn (2007), for example, considers natural laws as concrete things that must be explained in order to provide a complete explanation of all concrete existence.

should be rejected because both explanatory paths between concrete objects and natural laws seem to be metaphysical explanations. As indicated before, to provide a complete explanation of all concrete existence by natural laws, those laws must be considered as causally inert, abstract entities. So, one cannot account for the existence of all concrete things in terms of the causal force of prior laws; that explanation cannot be scientific. If that is true, then only the explanation of concrete objects by abstract natural laws can proceed successfully, but not the other way around. In the end, the obtaining of natural laws still remains unexplained, and the claims for scientific self-explanation are not justified.

III. SELF-SUBSUMPTION

Now, is it possible at all to provide an ultimate self-explanatory account of existence? In criticizing Loewer's suggested ultimate explanation, Lange (2013) speaks of a "general prohibition against self-explanation." He uses self-explanation not in the loose sense that something's essence explains its existence, but in the strict sense that some fact literally explains itself. To think that at least nothing contingent can explain itself is presupposed by many others (Brenner 2020, Holt 2012, O'Connor 2008, Parfit 1998, Pruss 2006, Swinburne 2004, Vintiadis 2018). While Hempel and Oppenheim (1965) argue that scientific explanations cannot be circular or self-explanatory, nevertheless, the presupposition that *no kind* of logically contingent fact can explain itself has, as far as I know, never been backed up by reason.

On the contrary, Nozick (1981) mentions a way that some law or principle may explain itself by being an instance of itself. He calls such a way of explaining "self-subsumption" and explicates it as follows: A principle that asserts "All principles of a certain kind are true" subsumes itself if it is a principle of that kind. Self-subsumption operates in the same way that the sentence "Every sentence of exactly eight words is true" is an instance of itself, except that the latter is obviously false. Nozick's suggestion for a valid self-subsuming principle is, instead, "All possible worlds obtain." Since the obtaining of all possible worlds is itself a possibility, that principle can be said to subsume itself. Nozick concedes, however, that self-subsumption cannot operate as proof or justification for truth. His claim is, rather, that *if* the self-subsuming principle is true, its obtaining can be explained in terms of itself. Although Nozick does not rule out the possibility that a self-subsuming principle itself is explained further, he offers self-subsumption as one possible way that one might be able to ultimately solve the Why question.

According to Wedin (1985), four phases are involved in Nozick's suggestion to answer the Why question by self-subsumption. First, Nozick attacks the presupposition that nothingness is a natural state and would need no explanation if

it obtained. Thus, he favours *Egalitarian* hypotheses, in which no state of affairs is arbitrarily considered as without a need to be explained. To avoid considering nothingness or any other state as natural, all of those states of affairs must obtain. The second phase is therefore *Fecundity* – the hypothesis that all possible worlds obtain. Third, Nozick attempts to explain the Fecundity hypothesis through self-subsumption. Finally, he upgrades the Fecundity hypothesis in order to avoid some difficulties. Most responses to Nozick's suggested ultimate explanation are critical. Many of them blame him for the obscurity of his suggestion. I attempt, however, to make clear what Nozick really intends and to defend his proposal of an ultimate explanation for all (concrete) existence.

Among the first critics, Wedin (1985) objects that Nozick uses "possibility" and "possible world" interchangeably when he maintains, as a self-subsuming case, that the obtaining of all possible worlds is itself a possibility. Wedin makes clear that possible worlds are not the same as possibilities because possibilities might be contradictory and unable to obtain all in the same realm. For this reason, one must regard Fecundity as different from the principle of Plentitude, according to which the maximal sum of non-contradictory possibilities obtains (cf. Lovejoy 1936/1964). In light of Wedin's criticism, one should consider the obtaining of all possible worlds in independent, non-interacting realms. Nozick himself emphasizes that his suggestion resembles Lewis's Modal Realism, in which all possible worlds concretely and independently exist (Lewis 1986). Still, Witherall (2017) objects that Fecundity leads to a contradiction: The possibility must also obtain that *not all possibilities are realized*, and the latter contradicts Fecundity itself.

In response, however, one should notice that a Fecundity hypothesis may surpass Modal Realism by maintaining realities constituted by worlds: In possible worlds, possibilities are realized, and, in possible realities, worlds exist. So, the possibility that *all possibilities are realized* can obtain in a reality that is independent of a reality in which *not all possible worlds are realized*. All possible worlds therefore can obtain, and no contradiction seems to occur. A reality is regarded here as a logical space or a set containing various combinations of possible worlds. At a concrete level, we have the existence of every possible world, as Modal Realism suggests. However, at a more fundamental level, there exists every possible set of those worlds, including one that is a set devoid of any possible world – an empty set. Consequently, the objection cannot be revived by asserting that some possible realities do not exist.[12]

More serious criticisms attack the explanatory role of Fecundity. Wedin (1985), Kusch (1990), and Lacey (2014) object that Fecundity only mentions possible ways that a situation might have turned out and claims that those ways obtain as well. However, that suggestion is, so they criticize, not enough to reduce the mystery of why that situation and those alternative ways are realized.

[12] Thanks to Daniel Kodaj for bringing this potential objection to my attention.

Consider, for example, the question: Why did you go to the gym rather than not? Fecundity seems to say only that I went to the gym in this possible world but not in another, which is a funny answer to a why question but not obviously an explanation. Nevertheless, I argue that the Fecundity hypothesis can preserve its explanatory role by reducing the arbitrariness of concrete existence. Less arbitrariness for a hypothesis is another theoretical virtue, which makes that hypothesis to be regarded, methodologically, as more likely to be true. While Unger (1984) explicitly maintains that all possible worlds must obtain because of minimizing arbitrariness, many use the same way of reasoning in theological and scientific contexts. To reduce arbitrariness, Swinburne (1996) holds that a being with an abundance of personal virtues must be actual, and Krauss (2012) argues for a plurality of physical universes with different natural laws (a kind of Multiverse). As a result, it seems rational, as Parfit (1998) argues, to accept that if Fecundity is realized, the explaining factor is its abstract feature of being less arbitrary. This fact averts another objection that claims: Fecundity may only be a universal generalization that happens to be the case, and if that is true, Fecundity cannot rule over itself as a law (Joseph Smith 1988). If Fecundity rules over all concrete existence, having that special feature is extremely likely to be the result of a fundamental law or principle that is metaphysically necessary, rather than simply happening to obtain. As another related objection, Greene (2011) considers Nozick's Fecundity hypothesis to be ad-hoc and unfalsifiable. I argue, however, that we have good reasons to regard the principle of Fecundity as a law with explanatory force. In the subsumption of the principle of Fecundity by itself, the subsuming principle, which obtains in a reality including worlds, can be deeper than the subsumed principle in a possible world. The Fecundity principle, therefore, satisfies the characteristics of a valid explanation both for concrete existence and for itself.

Nonetheless, Fecundity shares certain difficulties with Modal Realism. All of us have the fundamental conviction that the future resembles, at least to some extent, the past. Although it is logically possible that we turn, in a second, into a cabbage or to disappear, neither Fecundity nor Modal Realism can explain why one should expect nature to follow some regularities. Leslie (2014) argues that accepting the Fecundity hypothesis ruins our inductive inferences. Against Hume's objection to Induction, there is a general agreement that our world obeys, or at least behaves in a way that appears to obey, relatively simple laws, and the ultimate explanation of existence must explain why this is the case. Nozick (1981) himself concedes that we seem to be living in a world that appears more unified than what is required for us to originate and continue to exist in it. Fecundity therefore requires upgrading.[13]

[13] As an anonymous referee of this journal mentions, the Fecundity of all possible worlds and its self-subsumption resembles a theistic hypothesis developed by Aranyosi (2013) and

Nozick's proposal becomes very obscure and complex when it reaches the fourth stage. Wedin (1985) thinks that Nozick attempts to upgrade the principle of Fecundity by appealing to mystical experience, and that is his main criticism against Nozick. To ground the structure of all possibilities, Nozick argues that a third category must be real that involve both existence and non-existence. He adds then that only by personal mystical experience can one justify the reality of things that "nonexist". However, I argue that none of these complications is required if one considers the question "Why is there something rather than nothing?" in a limited sense to ask why there is anything *concrete*. Instead of considering things that nonexist (rather than things that simply do not exist), one can consider the category of non-concrete things simply as abstract realities. In the same way that non-spatial things cannot be coloured or uncoloured because the category of having colour does not apply to them, abstract realities do not exist, nor do they nonexist. The category of concrete existence does not apply to abstract realities.

IV. AN AXIOLOGICAL SELF-SUBSUMPTION

All these complications aside, I agree with Lacey (2014) in thinking that, to upgrade Fecundity, Nozick seeks a limited version of Fecundity that not only subsumes itself, but also accompanies us in our conviction of inductive reasoning. However, Nozick does not suggest what feature a limited Fecundity must have in order to satisfy the latter conditions. Besides, he mentions a further complication in case one finds such a limited Fecundity: The question still remains unanswered of why that limited Fecundity with its special feature obtains rather than another limited Fecundity. Lacey (2014) objects here to the extreme obscurity of Nozick in suggesting a limited Fecundity that outweighs others. But it seems that Nozick is disappointed after finding an ultimate explanation for all existence that prompts no further why questions. If there is an ultimate explanation, the question still remains unanswered, according to Nozick, of why everything is explained. It might seem without explanation, a brute fact, that everything has an explanation.

---

independently by Nagasawa (2016) to view God as the totality of all possible worlds. They reformulate the Ontological Argument for the existence of God to argue that a reality consisting of all possible worlds is that than which nothing greater can be conceived. However, the mere fact that the Ontological Argument can be employed in various ways to argue for the existence of different beings provides enough support that none of them can be used as proof. Other evidence needs to support a valid form of the Ontological Argument, and the evidence from the uniformity of nature and the reliability of our inductive inferences show that our world is more unique to be considered as one possibility that exists among all possible worlds that exist.

Nevertheless, it is better to postpone the difficulty of selecting one among several limited principles of Fecundity for when one finds some such principles. No such principle has yet been proposed, even by Nozick himself. I propose here a self-subsuming limited principle of Fecundity. I base such an account on the Axiological explanation of existence, the principle that says the existing world is required to exist because of its Goodness. I borrow a limited principle of Fecundity from Leslie's Spinozistic view, which claims that all Good possible worlds are required to exist (Leslie 1979). The self-subsumption is then borrowed from Rescher's justification for the Axiological explanation, the claim that the best possible world must obtain because it itself is for the best (Rescher 1984). Finally, I construct a self-subsuming limited principle of Fecundity by arguing that all Good possible worlds are required to exist. A possibility according to which all Good possible worlds exist is itself a Good possibility (as a reality), and, in turn, is required to exist. The irony is that Leslie himself does not agree with Rescher's self-subsumption, and Rescher does not accept Fecundity. In contrast, my suggestion for an Axiological ultimate explanation has both the elements of Fecundity and self-subsumption. As long as there are no other suggestions for a self-subsuming limited principle of Fecundity, one might find it reasonable to favour this proposal, if one finds it valid, in terms of its explanatory power. And if one day there will be another such suggestion, then the simpler hypothesis will be preferable.

Can there be a self-subsuming principle devoid of the latter defect? I suggest there is one. Consider the principle: All valuable possible worlds exist. It subsumes itself because the existence of all valuable possible worlds is itself a valuable possibility; the latter must therefore be actual according to the same principle. This suggestion conforms, in a way, to the Axiological explanation. Suppose we have enough reason for accepting that facts of value might explain existence, there may be a force in the nature of good things to exist. Why then shouldn't we accept that all good things might exist in separate worlds? The existence of many Good worlds seems to outweigh in value the state of there being only one Good world. If one accepts the existence of realities including co-existing worlds, one may rather deem the existence of all Good worlds as the best of all realities. If value rules over existence, then all valuable worlds are separately actual in a reality, a Meta-world.

One main objection to the idea of co-existing worlds is about the unifying factor that brings them together. However, there is no need here for a unifying factor other than value; all good worlds can exist through Axiological requiredness. Although these worlds are distinct, they are still unified by their value, which is the underlying reality that holds them together. This is similar to how different universes can coexist under the same laws of quantum physics, even though they may have their own separate spaces. While universes and good worlds are different, the Laws of Nature and Axiological requiredness are aware of both and responsible for their existence.

The latter helps us to avoid another objection: Should one not expect all good realities to exist as well? This leads to an explanatory regress, which is not satisfactory. But the objection can be avoided because the abstract facts of value can only bring about a unified reality with maximum value. And what maximizes the overall value of reality is the existence of all good worlds. So, if value rules over existence, no valuable world can be missed in reality. The axiologically unified combination of all good worlds constitutes the best reality.

A self-subsuming Axiological explanation can win over other explanations of existence that end in brute facts or explanatory regress. It has other advantages as well. If all valuable possible worlds exist, there is no need to point only to one best possible world. All candidates for the most valuable possible world can exist separately. Besides, without any further difficulty, one can accept that we live not in the best of all possible worlds, but in one among many Good possible worlds. Thus, the problem of evils in our world is simply resolved. In addition, since some level of orderliness is required for all valuable possible worlds, we can rely on our inductive inferences without finding afterwards that our world behaves differently from the past. The Axiological explanation can therefore be compatible with our conviction that the future resembles the past.

While Leslie is strongly against self-subsumption, his Axiological explanation results in something very close to the proposed Multi-worlds:

> …no possible existent would seem better than a mind worth calling "divine", a mind contemplating everything worth contemplating – this including, we might well think, every detail of possible universes in infinite number and endless variety… If the realm of existing things owes its reality to its creative ethical requiredness, then it must contain not just one infinite mind but infinitely many. Each contemplates absolutely everything worth contemplating (Leslie 2014).

If God is considered as "the power of being" or "the creative ground of existence," all the Multi-worlds, all valuable things, can be viewed as the creation of God. This makes a case for a Platonic theistic account of the world.

Yet, if one finally accepts a hypothesis as the ultimate explanation of all existence, some questions seem to remain unanswered: Why should one suppose that everything is explainable (unless one finds a reason that shows otherwise)? Why does one suppose that a hypothesis with more theoretical virtues is more likely to be true? Let's just accept for now that those are the presuppositions of reasoning without which one might not be able to know the world. It might not be even coherent at all to demand an explanation for the world's explicability. For, one who asks a why question already concedes that it is rational to demand explanations.

V. CONCLUSION

I argued in this article for an Axiological teleology, the claim that the world exists because it is intrinsically valuable. This hypothesis has the merit of answering a question that is not suitably answered by others. And that question is one of the most fundamental questions: Why is there anything concrete at all? To be sure, the world might simply have happened to exist. But if some fact explains why the whole world exists, that explanation must reside in the realm of abstract facts. No regress of concrete things whatsoever can explain why the whole concrete world exists. One main candidate for an abstract explanatory feature is the world's value. Only through the world's value can one provide an ultimate explanation for existence. However, value facts should not be regarded as necessary and not apt for further explanations. Just as many necessary facts have explanations, so too might facts of value be further explained. Facts of value, on the other hand, can be self-explanatory. The fact that valuable things exist is itself valuable. So, if our world exists because of its value, all other valuable worlds must exist separately. The Axiological explanation, so interpreted, does not fall into explanatory regress. By virtue of its explanatory power, therefore, one might prefer the Axiological explanation to other explanations of existence.

While scholars have generally doubted that self-subsuming principles can successfully provide explanations, I defended one such principle to ultimately explain all existence. After first outlining objections to the explanatory role of Fecundity, I clarified the extent to which it explains all concrete existence and also itself. First, I argued that the obtaining of all possible worlds can carry an explanatory role for concrete existence because it does not make the existence of a world arbitrary. Fecundity then was shown to subsume itself if one takes for granted realities including existing worlds. However, a destructive objection to Fecundity was that it cannot comply with the fundamental conviction that our world behaves in accordance with simple laws. An alternative self-subsuming principle is required, therefore, to explain why we live in a unified world of regularities. As a result, I further constructed a self-subsuming limited principle of Fecundity based both on Leslie's and Rescher's Axiological explanations for existence. My proposed ultimate explanation for all concrete existence is the principle that *all intrinsically valuable possible worlds are required to exist*. While that principle subsumes itself because it obtains in a valuable reality constituted by co-existing valuable worlds, it does not fail us in our conviction of inductive inferences. For, all valuable possible worlds behave lawfully and are unified. Still, some accepted features of our world remain unexplained; for example, the fact that the existence of every concrete thing is explained and that true hypotheses are simpler and have more explanatory power. One might accept these presuppositions in the end as the methodological rules of our reasoning. Or one

may maintain that it would be too late for demanding an explanation for the rationality of everything. For, one first concedes to the intelligibility of the world when one asks: *Why* is the existing world intelligible?

REFERENCES

Aranyosi, István 2013. *God, Mind and Logical Space: A Revisionary Approach to Divinity*. New York, Springer. https://doi.org/10.1057/9781137280329

Bird, Alexander 2007. *Nature's Metaphysics: Laws and Properties*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199227013.001.0001

Baldwin, Thomas 1996. There Might Be Nothing. *Analysis*. 56. 231–238. https://doi.org/10.1093/analys/56.4.231

Barnes, Eric 1994. Explaining Brute Facts. *PSA: Proceedings of the Philosophy of Science Association*. 1. 61–8. https://doi.org/10.1086/psaprocbienmeetp.1994.1.193011

Brenner, Andrew 2020. Explaining Why There Is Something Rather Than Nothing. *Erkenntnis*. 87. 1–17. https://doi.org/10.1007/s10670-020-00277-6

Cameron, Ross P. 2008a. Truthmakers and Ontological Commitment: Or How to Deal with Complex Objects and Mathematical Ontology without Getting into Trouble. *Philosophical Studies*. 140. 1–18. https://doi.org/10.1007/s11098-008-9223-3

Carroll, Sean M. 2018. Why Is There Something, Rather Than Nothing? https://arxiv.org/abs/1802.02231

Craig, William L. 2008. *Reasonable Faith: Christian Truth and Apologetics*. Wheaton, Crossway.

Della Rocca, Michael 2010. *PSR*. Ann Arbor/MI, Michigan Publishing.

Efird, David – Tom Stoneham 2005. The Subtraction Argument for Metaphysical Nihilism. *The Journal of Philosophy*. 102. 269–280. https://doi.org/10.5840/jphil2005102614

Fine, Kit 2012. Guide to Ground. In Fabrice Correia – Benjamin Schneider (Ed.) *Metaphysical Grounding: Understanding the Structure of Reality*. New York, Cambridge University Press. 37–80. https://doi.org/10.1017/cbo9781139149136.002

Goldschmidt, Tyron 2011. The New Cosmological Argument: O'Connor on Ultimate Explanation. *Philosophia*. 39. 267–288. https://doi.org/10.1007/s11406-010-9282-5

Gott III, J. Richard – Li-Xin Li 1998. Can the Universe Create Itself? *Physical Review D*. 58. https://doi.org/10.1103/physrevd.58.023501

Greene, Brian 2011. *The Hidden Reality: Parallel Universes and the Deep Laws of the Cosmos*. New York, Vintage.

Hawking, Stephen – Leonard Mlodinow 2010. *The Grand Design*. New York, Random House Digital.

Heil, John 2018. Must There Be Brute Facts? In Elly Vintiadis – Constantinos Mekios (Ed.) *Brute Facts*. Oxford, Oxford University Press. 19–27. https://doi.org/10.1093/oso/9780198758600.001.0001

Heller, Michael 2009. *Ultimate Explanations of the Universe*. New York, Springer Science & Business Media. https://doi.org/10.1007/978-3-642-02103-9

Hempel, Carl – Oppenheim, Paul 1965. The Logic of Scientific Explanation. In Carl Hempel. *Aspects of Scientific Explanation*. New York, The Free Press. 297–330. https://doi.org/10.2307/2272422

Holt, Jim 2012. *Why Does the World Exist? An Existential Detective Story*. New York, Liveright Publishing Corporation.

Krauss, Lawrence M. 2012. *A Universe from Nothing: Why There is Something Rather than Nothing*. London, Simon and Schuster.

Kuhn, Robert L. 2007. Why This Universe? Toward a Taxonomy of Possible Explanations. *Skeptic*. 13. 28–40.

Kusch, Martin 1990. On "Why Is There Something Rather Than Nothing?" *American Philosophical Quarterly*. 27. 253–257.

Lacey, Alan 2014. *Robert Nozick*. London, Routledge. https://doi.org/10.4324/9781315710785

Lange, Marc 2013. Grounding, Scientific Explanation, and Humean Laws. *Philosophical Studies*. 164. 255–261. https://doi.org/10.1007/s11098-012-0001-x

Lange, Marc 2014. Are Some Things Naturally Necessary? In Tyron Goldschmidt (Ed.) *The Puzzle of Existence: Why is there Something Rather than Nothing?* New York, Routledge. 235–251. https://doi.org/10.4324/9780203104323-18

Lange, Marc 2018. Transitivity, Self-Explanation, and the Explanatory Circularity Argument against Humean Accounts of Natural Law. *Synthese*. 195. 1337–1353. https://doi.org/10.1007/s11229-016-1274-y

Leibniz, Gottfried W. 1714/1989. Principles of Nature and Grace. In his *Philosophical Papers and Letters* (ed. Leroy E. Loemker). Amsterdam, Kluwer. 636–642. https://doi.org/10.1007/978-94-010-1426-7_67

Leslie, John 1979. *Value and Existence*. Oxford, Basil Blackwell.

Leslie, John 2014. A Proof of God's Reality. In his *The Puzzle of Existence*. London, Routledge. 136–151. https://doi.org/10.4324/9780203104323-12

Lewis, David 1986. *On the Plurality of Worlds*. Oxford, Basil Blackwell.

Loewer, Barry 2012. Two Accounts of Laws and Time. *Philosophical Studies*. 160. 115–137. https://doi.org/10.1007/s11098-012-9911-x

Lovejoy, Arthur O. 1936/1964. *The Great Chain of Being*. Cambridge/MA, Harvard University Press.

Mackie, John L. 1982. *The Miracle of Theism: Arguments for and against the Existence of God*. Oxford, Clarendon Press.

Mancosu, Paolo 2001. Mathematical Explanation: Problems and Prospects. *Topoi*. 20. 97–117. https://doi.org/10.1023/A:1010621314372

Maudlin, Tim 2007. *The Metaphysics within Physics*. New York, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199218219.001.0001

McDaniel, Kris 2007. Extended Simples. *Philosophical Studies*. 133. 131–141. https://doi.org/10.1007/s11098-006-9010-y

Moghri, Mohsen 2021. Deriving Actuality from Possibility. *Philosophia*. 49. 393–402. https://doi.org/10.1007/s11406-020-00246-x

Murphy, Mark 2011. *God and Moral Law*. Oxford, Oxford University Press.

Nagasawa, Yujin – Andrei Buckareff 2016. Modal Panentheism. In their *Alternative Concepts of God: Essays on the Metaphysics of the Divine*. Oxford, Oxford University Press. 91–105. https://doi.org/10.1093/acprof:oso/9780198722250.003.0006

Nozick, Robert 1981. *Philosophical Explanations*. Cambridge/MA, Harvard University Press.

O'Connor, Timothy 2008. *Theism and Ultimate Explanation: The Necessary Shape of Contingency*. Oxford, Blackwell. https://doi.org/10.1002/9781444345490

Parfit, Derek 1998. Why Anything? Why This? *The London Review of Books*. 20. 24–27.

Plantinga, Alvin 1977. *God, Freedom, and Evil*. Grand Rapids, William B. Eerdmans Publishing.

Plantinga, Alvin 2011. *Where the Conflict Really Lies: Science, Religion, and Naturalism*. New York, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199812097.001.0001

Pruss, Alexander 2006. *The Principle of Sufficient Reason: A Reassessment*. Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9780511498992

Rescher, Nicholas 1984. *The Riddle of Existence: An Essay in Idealistic Metaphysics*. Lanham, University Press of America.

Rescher, Nicholas 2010. *Axiogenesis: An Essay in Metaphysical Optimalism*. Lanham, Lexington Books.
Rescher, Nicholas 2013. *On Explaining Existence*. Berlin, Walter de Gruyter. https://doi.org/10.1515/9783110320442
Rodriquez-Pereyra, Gonzalo 1997. There Might Be Nothing: The Subtraction Argument Improved. *Analysis*. 57. 159–166. https://doi.org/10.1093/analys/57.3.159
Rosen, Gideon 2010. Metaphysical Dependence: Grounding and Reduction. In Bob Hale – Aviv Hoffmann (Ed.) *Modality: Metaphysics, Logic, and Epistemology*. New York, Oxford University Press. 109–135. https://doi.org/10.1093/acprof:oso/9780199565818.003.0007
Rowe, William L. 1970. Two Criticisms of the Cosmological Argument. *The Monist*. 51. 441–459. https://doi.org/10.5840/monist197054329
Russell, Bertrand – Copleston, Frederick C. 1948. *A Debate on the Existence of God*. New York, Harper & Row.
Smith, Joseph W. 1988. *Essays on Ultimate Questions*. Beatty, Avebury.
Smith, Quentin 1988. The Uncaused Beginning of the Universe. *Philosophy of Science*. 55. 39–57. https://doi.org/10.1086/289415
Swinburne, Richard 1997. *Simplicity as Evidence of Truth*. Milwaukee, Marquette University Press.
Swinburne, Richard 2004. *The Existence of God*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199271672.001.0001
Tillich, Paul 1951. *Systematic Theology*. Chicago, The University of Chicago Press.
Unger, Peter 1984. Minimizing Arbitrariness: Toward a Metaphysics of Infinitely Many Isolated Concrete Worlds. *Midwest Studies in Philosophy*. 9. 29–51. https://doi.org/10.1111/j.1475-4975.1984.tb00051.x
Van Cleve, James 2018. Brute Necessity and the Mind–Body Problem. In Elly Vintiadis – Constantinos Mekios (Ed.) *Brute Facts*. Oxford, Oxford University Press. 63–96. https://doi.org/10.1093/oso/9780198758600.003.0005
van Inwagen, Peter 1983. *An Essay on Free Will*. Oxford, Oxford University Press.
Vilenkin, Alexander 2007. *Many Worlds in One: The Search for Other Universes*. New York, Hill and Wang.
Vintiadis, Elly 2018. There Is Nothing (Really) Wrong with Emergent Brute Facts. In Elly Vintiadis – Constantinos Mekios (Ed.) *Brute Facts*. Oxford, Oxford University Press. 197–212. https://doi.org/10.1093/oso/9780198758600.003.0011
Wedin, Michael V. 1985. Nozick on Explaining Nothing. *Philosophy Research Archive*. 10. 337–346. https://doi.org/10.5840/pra1984107
Witherall, Arthur 2017. *The Problem of Existence*. London, Routledge. https://doi.org/10.4324/9781315237602

László Bernáth

The Aporia of Categorical Obligations and an Augustinian Teleological Way Out of It*

I. INTRODUCTION

Starting from the 1950s, three traditions emerged in analytic philosophy, which, among other things, focus on the critical examination of categorical obligations. One of them is the analytic revival of virtue ethics. Some proponents of this new wave of virtue ethics (most famously, Anscombe 1958 and MacIntyre 1985) have claimed that it makes no sense to build morality upon the notion of categorical obligations because this notion is unintelligible without a moral system that is forgotten in the modernity. Another, more recent tradition also criticizes the notion of categorical obligations. The proponents of moral error theory (or normative eliminativism) argue that because categorical obligations and the normative properties that are too closely related to them are queer from a physicalist/naturalistic perspective, it is highly implausible to suppose that such obligations and properties exist (Husi 2013; Olson 2014; Cowie 2016; Streumer 2017; Cote-Bouchard 2017). The last tradition that I would like to mention is, in fact, not a tradition proper, but the influential work of Bernard Williams, whose moral philosophy is hard to categorize because he advocates some kind of anti-theoretical attitude toward ethics. Regardless of whether one labels his work "Nietzscheian" or with any other fancy name, he certainly believed that the notion of categorical obligation is not only harmful to personal integrity, but also nonsensical because every reason has to be internal and somehow connected to the agent's motivations (see especially Williams 1981. 101–113.).

My problem with these challenges to categorical obligations is that in my view, the arguments in favor of them are either unnecessarily convoluted (I regard MacIntyre 1985 as such an example), or rely too heavily on some kind of

* The research was supported by the OTKA (Hungarian Scientific Research Fund by the National Research Development and Innovation Office) Postdoctoral Excellence Programme (grant no. PD131998), and another OTKA research grant (grant no. K132911). The paper is a vastly modified version of my talk entitled "The right to believe in the teleology of man" that I give online at the "New work on the metaphysics of teleology" webinar. I would like to thank for Boldizsár Eszes, Dániel Kodaj, Paár Tamás and an anonymous reviewer for their insightful comments on the talk and/or a previous version of this paper.

worldview. In the case of some virtue ethicists (most notably Anscombe 1958), a Divine Command Theory about categorical obligations seems to lurk in the background. In contrast, contemporary error theorists and their arguments are heavily influenced by physicalist/naturalist ontological assumptions – it is not a coincidence that they so frequently mention the alleged ontological queerness of moral properties and facts. In proposing that the main point of talking about reasons is to *explain* actions, even Williams himself reveals the naturalist underpinnings of his philosophy.

The reason why I bring up the influence of supernatural and naturalist views on these ethical approaches is not because I suspect that they distort their criticism of categorical obligations, but because I believe that the challenge to categorical obligations is rather simple and independent of ontological and ethical frameworks.[1] And I think that it is rather important that this simple problem with categorical obligations arises not because one has this or that worldview (and the previous accounts may give the impression that it can be the case) but because the concepts of 'reason', 'obligations', 'rationality' and 'motivational states' are connected to each other in a way that it is hard to make sense of the notions of 'categorical reason' and 'categorical obligation'. To see the gist of a conceptual problem clearly has its own value in itself, but I hope that my characterization of the problem can show why a specific – Augustinian – solution of the problem is the best if one does not takes its metaphysical price into account.[2]

In the second section of the present paper, I outline the key concepts in a way that is helpful in posing a challenge to categorical obligations. In the third section, I use this framework and describe the aporetic challenge to categorical obligations. In the fourth section, I argue in general terms that there is a promising way out of the aporia if one accepts that not only categorical but also quasi-categorical obligations can do the job that is needed in order to have a moral system with strong normative power. In the final section, I give a general outline of a system of quasi-categorical obligations that is based on the Augustinian view of motivations.

[1] To my mind, the best and most worldview-independent formulation of the problem can be found in Anthony Robert Booth's recent paper (Both 2022). However, it is swamped with a punctual but pedantic jargon that is necessary for closely engaging with recent debates on the subject. My goal is not to defend error theory against every recent objection (as Booth attempts to do) but to outline the problem of categorical obligations *both* accurately and *simply*.

[2] An anonymous reviewer objected that it makes not much sense that I attempt to reconstruct the problem in a worldview-independent way if the solution implies some kind of metaphysical worldview. I beg to differ. If one has a formulation of the problem that can be accepted by every rational person regardless of their worldview, it is a great step forward because in this case they can agree that the problem is there and it is not generated only by a part of this or that worldview. Moreover, one can more clearly compare the possible solutions to the problem even if all solutions imply some worldviews.

II. THE NOTIONS OF HYPOTHETICAL AND CATEGORICAL REASONS AND THEIR RELATION TO CATEGORICAL OBLIGATIONS

First, I would like to clarify the notion of categorical reason. Richard Rowland gives an elegant description of its content:

> Categorical normative reasons are normative reasons for agents to do things or have certain attitudes irrespective of their desires, aims, wants and feelings, and the roles in which they happen to find themselves; these reasons for agents to do things are ontologically/existentially independent of these agents' desires, aims, wants, feelings and roles. In contrast, hypothetical normative reasons are reasons for agents to do things or have certain attitudes that are not independent of these agents' desires, aims or roles. For instance, if, but only if, you like blueberry muffins, there is a reason for you to buy some. (Rowland 2013. 3)

Note that in itself, this is a rather formal definition of categorical reasons because only one difference between categorical and hypothetical reasons follows from it *logically*. This difference is explicitly mentioned in the definition: categorical reasons provide a reason to act regardless of your mental states or societal roles. Strictly speaking, it does not follow from this definition that categorical reasons have stronger normative force than hypothetical ones. That is, it does not follow from it that categorical reasons necessarily outweigh hypothetical ones and provide stronger reasons to act than "If…, then…"-type reasons. For instance, it could be the case that your hypothetical reason to steal some blueberry muffins outweighs your categorical reason to not steal them because the definitions do not exclude the possibility that you like blueberry muffins so much that this fact gives rise to a super-powerful reason to neglect your categorical reason to not steal them.

Of course, this idea seems to be absurd. That a desire for blueberry muffins can override a categorical reason contradicts any interpretation of the notion of categorical reason. This is not only because the idea is implausible but because this possibility makes the notion of categorical reason useless or even empty. Insofar as a hypothetical reason can be more powerful than a categorical reason, acting upon a categorical reason can be rational only if there is no such powerful hypothetical reason in the situation. However, whether there is such a powerful hypothetical reason in the practical situation depends on what your desires/aims/wants/feelings are. That is, if hypothetical reasons could be more powerful than categorical reasons, then whether acting upon a categorical reason is rational would depend on what your desires/aims/wants/feelings are. So, in this case, even if *the existence of categorical reasons* did not depend on what your motivational states are, the *rationality* of acting upon categorical reasons in any specific situation would depend on these states.

The problem is that if (i) the rationality of acting upon categorical reasons depends on the mental states of the agent and (ii) one should do in any situation what is among the most rational options, then categorical obligations binding agents irrespectively of their desires/aims/wants/feelings are impossible. Let us suppose that John likes blueberry muffins so much that in a concrete situation the hypothetical reason to steal them outweighs his categorical reason to not do that. Now, if John did not like blueberry muffins, would the categorical reason outweigh any other reason, and would it be rational for him to act upon it? Not necessarily, if he has another powerful hypothetical reason to steal the blueberry muffins. If, let us say, he just liked to steal stuff very much, then it would still be rational for him to act against his categorical reason to not steal anything. Insofar as hypothetical reasons can, in principle, outweigh categorical reasons, it would be rational for John to act upon his categorical reason to not steal anything only if he cared enough about what his categorical reasons are. Moreover, since one should act rationally, one can say that John should act upon his categorical reason to not steal anything only if he cared enough about what his categorical reasons are. Thus, if categorical reasons can be outweighed by hypothetical reasons, then it is rational for John (or anyone) to transform the categorical reasons into the form of this hypothetical obligation: "I should act upon *R* only if I care enough, relatively to my hypothetical reasons, about my categorical reason *R*." So, in this case, it would be possible for everyone to derive only hypothetical obligations from categorical reasons every time and everywhere, and categorical obligations could not exist at all. This is because categorical obligations would be precisely those obligations upon which one should act regardless of one's desires, aims, wants and feelings, and the roles in which one happens to find oneself.

Consequently, categorical obligations are possible only if there are some categorical reasons that cannot be outweighed by any hypothetical reasons. The situation, however, seems to be aporetic. One could block the transformation of categorical reasons into hypothetical obligations only in two ways. Firstly, one could deny that agents should do the most rational option or at least one of the most rational options. I do not embrace this possibility because I assume that most people would agree with this claim, and I think that practical rationality should be defined through the notion of "should". Here is one such definition. A reason is something that you can use in a reasoning to justify an action. To justify an action is to show either that you are permitted to do the action (if not doing the action can be justified, too) or that you should do the action (if not doing the action cannot be justified).[3] Secondly, one can deny that the rationa-

[3] This picture is similar to but much simpler than Derek Parfit's approach toward reasons and obligations (see Parfit 2011. 31–37). In this context, the notions of hypothetical and categorical obligations are more useful than the difference between normative and motivating

lity of acting upon any reasons depends on the agent's mental states. In the next section, I will argue that this method for blocking the transformation of categorical reasons into hypothetical obligations does not work because mere taboos do not provide any reasons, and other taboo-like reasons are, in fact, masked hypothetical reasons.

III. THE IMPOSSIBILITY OF NON-TRANSFORMABLE CATEGORICAL REASONS AND THE APORIA OF CATEGORICAL OBLIGATIONS

To see why categorical reasons that are non-transformable into hypothetical obligations are impossible, the notion of taboo is a useful starting point. This is because taboos are those imperatives that most clearly have the form of categoricity, but it is easy to see that they cannot be transformed into categorical obligations. MacIntyre describes the notion of taboo in the following way:

> Captain Cook and his sailors were told [in the Pacific Islands] that men and women could not eat together, because it was *taboo*. But when they enquired what that meant, they could learn nothing except that it was an absolute and unconditional requirement which could not be further explained. We do not take *taboo* seriously; why then should we take seriously Kant's or Prichard's *ought?*[4] (MacIntyre 1981. 124–125)

---

reasons. The argument in the next section can be seen as one for the claim that every normative reason (reasons which *really* count in favor of doing something) is hypothetical at the end of the day because every normative reason can count in favor of doing something only if it has the appropriate relation to an actual motivational state.

[4] In this relatively early text, MacIntyre argues for a similar conclusion as I do in this paper. His critique of the theories of categorical reason is similar to mine. Nevertheless, I think that my argumentation is much more compact and clearer because, contrary to MacIntyre, I outline the nature of the relation between reasons and obligations. However, the main advantage of my approach can be seen in the next section in which I turn to the modal aspect of the problem. The proper differentiation between reasons and obligations opens up the possibility of quasi-categorical obligation that is a much better solution for the problem of categorical obligation than MacIntyre's early theory, in which he claims that categorical obligations arise when one is part of a wider narrative. The problem with this solution is that one can choose one's wider narrative in which one takes part; therefore, categorical obligations still do not necessarily bind the agent. An additional problem is that this solution seems to imply moral relativism in the long run. MacIntyre's early narrative theory, in a less bold form, seems to be a part and parcel of *After Virtue* as well. It is important to note that there are some who argue that this kind of narrative theory does not imply moral relativism (for example: Kuna 2005; Renani 2017). Also, it is worth mentioning that MacIntyre has developed his theory of narrativity into a more metaphysical approach that is similar to the supernaturalistic conclusion of this paper (see, for instance MacIntyre 2017. 52–59; 228–231; 314–315). Interestingly, he does not closely connect this new theory with the problem of categorical obligation. Rather, he focuses on how one can interpret one's life through exercising practical rationality.

I think that this description is a rather good characterization of the problem of *mere taboo*. We do not, and, I would say, even cannot take a *mere taboo* seriously. This is because *mere taboos* do not provide any reason for acting, as it is totally unintelligible why it would be good to act upon them. Thus, they cannot provide any categorical reasons that can be transformed into categorical obligations (because they are not reasons at all). Captain Cook and his crew regarded the imperative for women and men to eat separately as a *mere taboo*, and for them, taboos like this were not reasons that could possibly guide their actions. However, the people of the Pacific Islands in one way or another do not regard this *taboo* as a *mere taboo*. This prescription was built into the very fabric of their culture and endorsed by the authorities whose orders were relevant for them. So, members of the tribe see this prescription as a part of a reliable culture and as enjoined by a group of authorities; thus, they suppose that both their culture/authorities and the taboos serve the interests of the people. *Mere taboos* are unintelligible and do not provide any reasons at all; however, people who endorse *taboos* do not regard them as mere taboos, but rather as prescriptions that help them to achieve something valuable even if these taboos make it happen in an unknown way and the specific value of the taboos' fulfilling their purpose is not so well-defined.

The problem is that if we transform a mere taboo into a reason by adding either a clear or a vague goal to it, the result will be no more than a hypothetical reason. If the taboo serves the needs of, let us say, the people of the Pacific Islands, then the taboo gives only a hypothetical reason: "If you care enough about the needs of the people of the Pacific Island, it is reasonable for you to do *A*". It is not clear how we could transform a mere taboo into a reason in another way.

The issue is independent of the ontological status of the *taboo*. Let us suppose that there is a platonic state of affairs, namely that men and women could not eat together, a kind of platonic *taboo*.[5] Still, it does not qualify as a reason because it is unintelligible why one should act on the basis of this platonic state of affairs. Why is it better if this taboo is written in the sky rather than the dirt? Of course, if one likes to act in a way that one's actions fit this platonic state of affairs, then it is not a mere taboo any more, and, for one, the existence of this platonic fact serves as a reason, but it is still nothing more than a hypothetical reason. Such platonic entities can at best provide a hypothetical reason rather than a categorical one. Moreover, we are not better off if we refer to an Aristotelian fact that has something to do with objective teleology, which is independent from what our motivational states are. The mere fact, for example, that our body serves the goal of sustaining our life gives us a reason to not commit suicide only if we care about the existence of such an Aristotelian fact.

[5] This part of the text is inspired by Erik J. Wielenberg's metaethical views (especially Wielenberg 2009), but I do not pretend that I can refute his complex views in one paragraph. Nonetheless, I would go in a similar direction if my aim were to criticize him.

The problem does not relate to the question whether it is reasonable to accept a *mere taboo* as a law. It seems to me that it is reasonable for everyone to accept the imperative "Do not steal for fun!" as a law for the society because if too many people stole stuffs, no one's property would be in safety, not to mention the potential consequence that the institution of property would cease to exist.[6] Nevertheless, it does not follow from the foregoing that it is rational for the agent to act upon this accepted law. From the perspective of a clever but very selfish person, the most rational thing for them to do in order to serve their selfish needs is to publicly endorse the laws against stealing, but steal things whenever it fulfills their desires and there is only a negligible chance of getting caught. The mere fact that you have accepted a law or should accept a taboo as a law implies at best only that you have a hypothetical reason to act upon a taboo. This hypothetical reason goes like this: "If you care enough to act in accordance with the law, it is worthwhile for you to obey the law". The reason remains hypothetical even if you would like to follow the law not because you obey the law out of your pure respect of it, but because you believe that the law, in general, helps to actualize some values.

It seems that anything can be a reason for anyone to do *A* (or not to do *A*) only if it can, in principle, make it intelligible why doing *A* is good in some way. However, if one regards doing *A* as worthwhile to do because it is good in some way, then if one does not care enough about the kind of good cited in our explanation, it will be rational for one to not try to achieve the kind of good in question. This means that we are in an *aporia*, because it seems that moral obligations are categorical obligations (they bind the agents regardless of what their motivations are), but reasons can only be hypothetical. What is more, hypothetical reasons cannot form the ground of categorical obligations. Or can they? Perhaps the way out of the aporia lies in this unexpected direction.

[6] I cannot hide the fact that this paragraph goes against the contractualist/constructivist tradition on morality. Once again, I know too well that it would be futile trying to refute such an influential tradition in one paragraph. It would be too lightweight even to nudge Kant's philosophy because Kantian constructivism (and other types of constructivism, too) has a specific view on practical reason; namely, that practical reason is, first and foremost, a law-making faculty. I do not try to argue against this approach because I have other goals, and I hope that this concept of practical reason is implausible for most readers. In my view, the (constructivist) theory that practical reason is primarily a law-making faculty is constructivist in the additional sense that it is a construction by philosophers trying to achieve philosophical goals.

IV. THE POINT OF CATEGORICAL OBLIGATIONS AND WHY (QUASI-)CATEGORICAL OBLIGATIONS CAN BE BASED ON HYPOTHETICAL REASONS

The aporia of categorical obligations emerges for two reasons. First, we rationally and consciously act always in order to achieve some goals, some values; thus, a reason for doing *A* has to refer to some value. Second, any reason for doing *A* which refers to some value as a desirable goal can have any normative power only if the agent cares enough about achieving the value in question. That is, all reasons have to be hypothetical.[7] And if a reason is hypothetical, it can be a source only of hypothetical obligations, because if it counts as a reason only if the agent has some specific mental states, then it cannot be a basis for an obligation to do *A* that binds agents irrespectively of their mental states.

At this point, it is worth asking why categorical obligations are so important anyway. The tripartite answer to this question is that firstly, they are important because insofar as there were only hypothetical obligations, we should confess that it is perfectly okay and rational for one to do the most abhorrent things if one's desires and goals are sufficiently twisted and one is powerful enough. If moral obligations, let us say, bind one only if one cares enough, then they will have no normative grip on those who just do not care. Secondly, if there are no categorical obligations, then moral progress is optional for the individual. If moral obligations are hypothetical, then to develop moral skills that make one able to fulfill them is just an option among many. If you are a liar who does not tell the truth in some situations, it is perfectly okay if you do not change, just as it is perfectly okay if you do not train yourself to be a better tennis player (see Wittgenstein 1965. 5). Thirdly, the non-existence of categorical obligations would make self-loathing perfectly rational for morally good people. This is because fulfilling moral obligations frequently includes self-sacrifice. However, self-sacrifice is painful, and if it is only an option among many, then morally good persons who are hypothetically obligated to make these painful sacrifices can reasonably say to themselves that it would be better if their goals and desires were different, because they would suffer much less.

All of these problems (and potential further ones) make it clear that categorical obligations are important because only they can bind agents *necessarily* (irrespectively of their desires, goals, etc.). So, if hypothetical reasons could necessarily bind agents, then they could give rise to categorical obligations or quasi-categorical obligations.

[7] In contrast with Williams's argumentation (Williams 1981), the outlined reasoning does not rely on the presupposition that the main function of reasons is to explain actions. Rather, it is based on the rather basic assumption of most theories of action that rational and conscious actions are intrinsically goal-directed.

I talk about quasi-categorical obligations because it is hard to see how obligations that are completely independent of the agents' mental states are possible (and the notion of categoricity is historically tainted with this aspect) if they are based on hypothetical reasons. It is much easier to show how hypothetical reasons can bind agents *necessarily*, irrespectively of *some features* of their *mental states*.

Let us focus on the modal aspect of the problem. "If you are a human, it is worthwhile for you to learn literature". This is a quasi-hypothetical reason because not all agents are human, but they are the only ones for whom it is good to learn literature (even though this reason has nothing to do with mental states). Nevertheless, it is a necessarily binding reason for each human agent if each of them is necessarily human (if, let us say, being human is an essential property of all humans). To put it in metaphysical jargon: If an agent *S* is human in all possible worlds, and if *S* has a reason to learn literature provided that *S* is a human, then *S* has a reason to learn literature in all possible words in which *S* exists; that is, *S* necessarily has a reason to learn literature. Thus, this case of a hypothetical reason shows that such a reason can *necessarily* be a reason for an agent if it offers a reason in relation to a necessary property of the agent.

Now, the problem is that morally relevant properties seem to be contingent in the above metaphysical sense, since all of the agents' desires, goals, feelings, aims, wants etc. seem to be contingent. I think that it is plausible to assume that agents necessarily have desires/goals/aims (even at those moments when they do not have phenomenal consciousness), but still, the desire for any particular object is contingent. This is a serious problem because, as I argued, our reasons for acting cannot be conceived without pointing toward something as a possible object of our desires/aims/goals etc. Each full and real (hypothetical) reason to act has to be formulated in a way like this: "If you desire enough to know literature, it is worthwhile for you to learn literature".

There is a strong argument for the case that no object of any desire/goal/aim can be a necessary object of these attitudes. This is because every possible object of our desires/goals/aims was rejected by someone somewhere. Most people desire to live, yet, some desire to not, most desire pleasure, yet, some desire pain, most desire good for their children, yet, some desire that horrible things happen to their children, most desire to go to heaven, yet, some prefer to go to hell instead. Presumably, if other people lack an object corresponding to that of my desire/aim/goal, it is not metaphysically necessary that my desire/aim/goal has this object.

V. THE AUGUSTINIAN MODEL OF MOTIVATIONS AND MORALITY

As far as I can tell, there is only one way to make sense of the claim that agents' desires/aims/goals are necessarily directed toward something. One has to say that agents have some desires/aims/goals that are directed toward more than

one thing. Of course, they are contingently directed toward some objects, but their direction transcends their objects and is necessarily aimed at an objective beyond them. I call this the Augustinian model of motivational states.[8]

The Augustinian model claims that we desire more than we are aware of.[9] All (or at least, some of) our desires (or other motivational mental states) have a dual structure. On the surface, one desires to achieve an object, say, one desires to make more money. However, it is a rather common experience that achieving the object of our desire does not satisfy the desire in question. It happens many times when agents achieve the object of their desire that the desire does not disappear and the promised happiness does not come. In many cases, there is even a bit of disappointment over the lack of perfect satisfaction. What is more, the lack of perfect satisfaction and this bit of disappointment give rise to a different object of desire, because the desire is still there. Nevertheless, changing the object of the desire – let us say, starting to desire for traveling rather than for making more money – does not solve the problem, and one cannot get perfect satisfaction by achieving the new goal. This is because the desire is directed not only toward its object but also to its objective. The objective has a far greater value than the object of the desire, and the Augustinian insists that this fact explains why achieving the object of the desire does not perfectly satisfy the agent.[10] Furthermore, even though such desires are metaphysically contingently directed toward their objects, they are metaphysically necessarily directed toward the objective that has more value from the agent's perspective than any object.

[8] I do not claim that no one before Augustine of Hippo held that view. Personally, I think that Plato entertained this picture of motivation in some of his dialogues (in my view, *Symposium* is one of them). Nonetheless, I believe that the most memorable description of the dual nature of our motivational system can be found in Augustine's *Confessions*.

[9] I believe that the unique feature of the Augustinian model compared to other teleological models is making the analysis of motivational states of the center of the investigation of human nature. This is not a stark contrast, but it is notable that the Aristotelian approach is more focused on the analysis of rationality and other abilities. If one accepts that there can be only hypothetical reasons for acting, then a motivation-centered approach comes handy for answering the challenge that is posed by the acceptance of this thesis. It is worth to note that, in my view, MacIntyre's approach in *After Virtue* is clearly an Aristotelian one due to its focus on rationality and (context-relevant) moral virtues whereas his later work, *Ethics in the Conflicts of Modernity: An Essay on Desire, Practical Reasoning and Narrative*, has much stronger Augustinian tendencies.

[10] As an anonymous reviewer of the paper pointed out, not the Augustinian explanation of the lack of perfect satisfaction is the only possible explanation. Another well-known tradition explains the lack of perfect satisfaction not by an objective of the desire but by the human nature which makes perfectly fulfilling its desires impossible (I think Schopenhauer can be interpreted in such a way). I agree with that there is a plurality of possible explanations in this regard. However, I investigate the possible explanations of the lack of perfect satisfaction from the perspective of solving the problem of the categorical obligations. And it seems to be the case, as far as I can tell, only the Augustinian explanation of this fact (or something very similar) provides an explanation that can help in solving the problem of categorical obligations.

In fact, such desires of all humans are *primarily* directed toward their objective *through* their objects that are not the ultimate goals of these desires, but attempts to approximate their ultimate goal.

For the purposes of the present paper, it is not relevant which motivational states are at the center of the Augustinian model (although I think that desires are the best candidates). Instead, what is important is that (a) (some of)[11] our *actual motivational states* are, in the metaphysical sense, necessarily directed toward this ultimate value and (b) we – as humans – necessarily have such motivational states; and, last but not least, (c) since the relevant motivational states are all directed toward this ultimate value, it cannot be the case that we as agents do not care enough about achieving this ultimate value. If one necessarily wants this ultimate value without the possibility of changing the objective/losing the motivational state all together/being overridden by another motivation, then this motivational mental state can form the basis of morality's special normative power. In this case, the hypothetical reason "If you care enough about the ultimate value, then you should do *A*" can be transformed into the quasi-categorical obligation "You should do *A* (regardless of what the objects of your motivations are)". It is only quasi-categorical because the fact that the relevant *motivational states* are directed toward an ultimate goal makes it possible for the hypothetical reason to generate an obligation which binds agents irrespectively of what their projects are. So, this obligation is not totally independent of motivational mental states. If it were, it would be a mere taboo, and it would instantly devolve into a hypothetical obligation. Thus, this solution finds a way out of the original aporia by denying the need for categorical reasons and obligations to ground morality, and by providing a combination of hypothetical reasons, quasi-categorical obligations and a philosophical anthropology that can make the existence of necessarily binding obligations intelligible.

The Augustinian thesis that humans have an essential property of desiring for some ultimate value has another advantage, namely that it can explain why one's trying to be a morally better person is not simply one of those difficult and admirable possible enterprises that are optional to pursue. This is because ever-

[11] The Augustinian does not need to claim that every motivational state has a dual structure. As far as I am concerned, it would be implausible to say that the urge to scratch my back has a dual structure because this kind of motivation can lead to sufficient satisfaction without giving rise to a new motivation. The urge to scratch does not even seem to be formed in a rational and conscious way to any extent; on the contrary, its origin can be found solely in the irrational unconscious region of my mind. In contrast with the case of the urge to scratch my back, if I consciously decide to aim at creating itching in order to get satisfaction from getting relief, then I have a desire for scratching my back, and this desire has a dual structure. Even if I successfully cause – somehow – the feeling of itching and scratch my back, the relief does not bring perfect satisfaction, and a new object of my desire emerges. Nonetheless, urges seem to be irrelevant with regard to the problem of categorical obligations, so the Augustinian can focus only on the relevant motivational states such as desires.

yone is condemned to pursue this ultimate value, and in part, *this is what makes everyone human*. Even if one does something that pushes achieving the ultimate value further away, one does so because one is acting upon a desire that is directed toward the same ultimate goal. In this Augustinian picture, morality is a guideline about how to achieve this ultimate goal toward which every relevant desire is directed, and *failing to achieve this ultimate goal is absolute failure* because it is the failure to achieve the objective of *every relevant motivational state*.

Of course, the above makes it intelligible why literal or non-literal self-sacrifice can be a moral obligation. There can be situations in which our attempts to achieve any actual objects of our desires take us further from the ultimate value, and in these situations, we must do what we need to do in order to get closer to the ultimate value even if it means that we have to sacrifice our way of life or, in extreme cases, our life itself. If the Augustinian picture is correct, the morally good persons' self-loathing is inappropriate and, in the final analysis, irrational because their sacrifice serves their ultimate need besides that of other people. It is true even if the need in question has a very different nature than all the other needs that are sacrificed in the act of a perfect self-sacrifice.

The desire for the ultimate value has a different nature than other desires for two interconnected reasons. Firstly, it is not simply an additional one to all the other desires. Rather, the desire for the ultimate value pervades every relevant desire, similarly to the way God is not a being among many beings but Being itself, in whom everything lives, moves and has its being. Secondly, whatever the ultimate value may be, it should be something that is greater than the content of any concept, because any possible content of our concepts can be an object of our desire, and it is plausible to assume that nothing can provide perfect satisfaction if it does not contain something more than the content of any of our concepts, and therefore is not directed toward the goal toward which every relevant desire of ours is ultimately directed. Thus, the ultimate value should be supernatural because no science or philosophy can grasp it perfectly. It follows that contrary to the aspects of object-directed desires, the aspect of all relevant desires directed toward the ultimate value is supernatural in this sense.

The above means that the Augustinian interpretation of morality has an ontological price. The Augustinian picture regards the moral system of quasi-categorical obligations as that of imperatives pointing out to agents with great normative force what they must do in order to gradually approximate the ultimate value. However, were the ultimate value not to exist, our motivational states could not ultimately be satisfied, and they would not be directed toward anything that goes beyond their objects. In this case, morality would not tell us how to approximate the ultimate value, which would remain unintelligible, and the Augustinian picture of morality would fail. Thus, those accepting the Augustinian interpretation of morality have to posit the existence of an ultimate and supernatural value.

I leave it to the reader to decide whether the Augustinian answer to the challenge to categorial obligations is a good one. I believe that whether one considers it appropriate depends on one's other philosophical convictions. Nevertheless, the cost-benefit analysis of this view is simple. On the one hand, as I already noted, it has a non-negligible metaphysical price. On the other hand, it draws on a plausible picture of practical reasoning and motivations to explain how morality can have unmatched normative power that necessarily binds each individual.

REFERENCES

Anscombe, Elizabeth 1958. Modern Moral Philosophy. *Philosophy*. 33. 1–19. https://doi.org/10.1017/S0031819100037943

Augustine of Hippo 397/1944. *The Confessions of St. Augustine*. Trans. F. J. Sheed. London, Sheed & Ward.

Booth, Anthony R. 2022. The Type-B Moral Error Theory. *Erkenntnis*. 87. 2181–2199. https://doi.org/10.1007/s10670-020-00297-2

Cote-Bouchard, Charles 2017. *Belief's Own Metaethics? A Case Against Epistemic Normativity* (Ph.D. Dissertation). King's College London.

Cowie, Christopher 2016. Good News For Moral Error Theorists: A Master Argument Against Companions in Guilt Strategies. *Australasian Journal of Philosophy*. 94/1. 115–130. https://doi.org/10.1080/00048402.2015.1026269

Husi, Stan 2013. Why Reasons Skepticism is not Self-Defeating. *European Journal of Philosophy*. 21/3. 424–449. https://doi.org/10.1111/j.1468-0378.2011.00454.x

Kuna, Martin 2005. MacIntyre on Tradition, Rationality, and Relativism. *Res Publica*. 11/3. 251–273. https://doi.org/10.1007/s11158-005-0558-8

MacIntyre, Alasdair 2016. *Ethics in the Conflicts of Modernity: An Essay on Desire, Practical Reasoning and Narrative*. Cambridge, Cambridge University Press.

MacIntyre, Alasdair 1985. *After Virtue* (2nd ed.). London, Duckworth.

MacIntyre, Alasdair 1981. Can Medicine Dispense with a Theological Perspective on Human Nature? In Daniel Callahan – H. Tristram Engelhardt (Eds.) *The Roots of Ethics: Science, Religion, and Values*. Boston/MA, Springer US. 119–137. https://doi.org/10.1007/978-1-4613-3303-6_6

Olson, Jonas 2014. *Moral Error Theory: History, Critique, Defence*. Oxford, Oxford University Press.

Parfit, Derek 2011. *On What Matters, Volume I*. Oxford, Oxford University Press.

Renani, Ali Abedi 2017. MacIntyre's Moral Theory and Moral Relativism. *Philosophical Readings*. 9/3. 171–174. https://doi.org/10.5281/zenodo.1067280

Rowland, Richard 2013. Moral Error Theory and the Argument from Epistemic Reasons. *Journal of Ethics and Social Philosophy*. 7/1. 1–24. https://doi.org/10.26556/jesp.v7i1.69

Streumer, Bart 2017. *Unbelievable Errors: An Error Theory about all Normative Judgements*. Oxford, Oxford University Press.

Wielenberg, Erik J., 2009. In Defense of Non-Natural, Non-Theistic Moral Realism. *Faith and Philosophy*. 26/1. 23–41. https://doi.org/10.5840/faithphil20092612

Williams, Bernard 1981. External and Internal Reasons. In his *Moral Luck*. Cambridge, Cambridge University Press.

Wittgenstein, Ludwig 1965. A Lecture on Ethics. *Philosophical Review*. 74/1. 3–12. https://doi.org/10.2307/2183526

Ferenc Huoranszki

Intentional Actions and Final Causes*

Davidson once asked what events in an agent's life are her deeds and doings in contrast to those which merely happen to her (Davidson 1971/1980. 43). Since agency, *prima facie* at least, is a causal concept, it seems natural to approach this question by trying to understand the nature of causation that is involved in exercising it when agents act intentionally. But many years later, Davidson reported that he remained convinced "that the concepts of event, cause, and intention are inadequate to account for intentional action" (Davidson 1987/2004. 106).

My purpose in this paper is to argue that Davidson is right if causes are understood as efficient rather than final causes. I shall argue that the intentionality of behavior is an irreducibly teleological phenomenon, and hence we cannot dismiss the idea of final causation in our account of intentional action. Intentional actions have, of course, efficient causes, and in certain contexts those causes can even explain what an agent did. But efficient causes cannot explain, as the still dominant causal theories of actions aim to do, why what the agent did was intentional.

According to the standard version of causal theories, actions are intentional if they are caused, 'in the right way', by an agent's psychological attitudes or by some agent-involving mental event.[1] According to another version of such theories, an agent's behavior is intentional to the extent that the results of her behavior are caused by the *agent* understood as a persisting substance rather than being caused by some of her psychological states or by some mental event.[2] Although the two sorts of theories differ both in their ontological presuppositions

* Many thanks to two anonymous referees for their supportive and helpful comments on an earlier version of this paper and to Dániel Kodaj for urging me to do something with it.

[1] Such accounts include, among others, Goldman 1970, Searle 1983, Bishop 1989, Mele 1992, Enc 2003. For further reference see Davies 2010.

[2] Taylor 1966; Alvarez and Hyman 1998; Lowe 2009, Steward 2012. According to what strikes me as a somewhat peculiar mixture of these two approaches, although actions in general might only have events as their efficient causes, an agent's *free* actions must be caused by the agent. See O'Connor 2000 and Clarke 2003.

and in their accounts of the intentionality of actions, they share the common assumption that intentional agency must be understood in terms of prior *efficient causes*.[3] And this is exactly what teleological accounts of actions deny.

Traditionally, most philosophers who reject the causal theories argue that explanations of intentional actions with reference to agents' reasons cannot be causal.[4] But this, in itself, even if right, does not prove that the nature of intentional action and agency can only be understood in terms of final causes. Elizabeth Anscombe has famously claimed that intentional actions "are actions to which a certain sense of the question 'Why' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting" (Anscombe 2000. 9). However, agents can act intentionally even when they have no reason for which they act because an action done without or against one's reason need not be aimless. And further, behavior can be intentional even in such cases when the question does not seem to be applicable at all, unless "giving a reason" is simply understood as a synonym of "ascribing a goal".

Hence, I shall argue that intentional behavior, irrespective of whether or not it is done for a reason, is irreducibly teleological. Agents act intentionally if their behavior has some 'final cause' in the sense that they have some end or goal for the sake of which their actions are performed. The distinction between those forms of behavior which reveal intentional agency and those which do not cannot be understood in terms of prior efficient causes. Neither event-causation nor agent-causation can explain the intentionality of behavior. Agency, to the extent it is manifested by intentional actions, is a fundamentally teleological concept.

I shall argue for this thesis as follows. First, I shall distinguish two questions about the intentionality of actions: one that is related to the teleological structure of behavior and one that is related to the possibility of conscious control. I shall argue that the former is the more fundamental. Second, I shall highlight those aspects of Aristotelian final causation that make it especially fit for explaining the nature of intentionality of behavior. Third, I shall argue that granting that agents as persisting substances can cause events is neither necessary nor sufficient for the explanation of the intentionality of behavior; and further, that the psychological or psychophysical *origin* of behavior cannot explain its intentionality either. Consequently, we cannot understand the intentionality of actions with reference to inner efficient causes. Finally, I shall argue that "trying to" expresses the intentional mode or aspect of agents' behavior precisely because behavior is intentional only if it is done for the sake of some ends.

[3] When I say "only", I do not mean, of course, that the disagreement is minor or irrelevant.

[4] See, among others, von Wright 1971, Wilson 1989, Ginet 1990, Sehon 2005, Lowe 2009, and McLaughin 2012. Thompson's 'naïve action theory' might be interpreted as a version of the teleological view, see Thompson 2008. On Anscombian teleological accounts see also Wiland 2012. 145–155.

I. TWO CONTRASTIVE QUESTIONS ABOUT THE INTENTIONALITY OF BEHAVIOR

Sometimes we wonder whether an agent's behavior was intentional rather than being *non*intentional. *Prima facie*, what we inquire in such cases is whether the agent performed the action with some purpose. When we understand agents as acting intentionally, we typically see their actions performed as a means for some end. At least, this is how we ordinarily make sense of intentional behavior. When we try to understand an action *qua* intentional, we search for some end for the sake of which it was performed.[5]

Some other times, however, when we ask whether an agent's behavior was intentional, we are interested in something else. We are interested in whether the action was intentional rather than being *un*intentional. What we wonder in this case is whether an agent has succeeded in exercising conscious control over what she has done or failed to do. Raising this question presupposes that the agent must have had something in mind by doing what she did; or that she acted in the way she did because she had the intention, or at least *an* intention, to perform an act.[6]

Most intentional human actions are intentional in both senses: they are instances of purposeful behavior which the agents whose actions they are also had in mind. We expect normal adult agents to exercise some degree of conscious control over their own behavior, which they can do only if they are aware of what they intend to do. However, as far as the philosophical problem of agency and intentional behavior is concerned, the interesting issue is the relation between these two features of intentional behavior: whether the possibility of purposiveness depends on the possibility of conscious control; or rather, whether conscious control presupposes the purposefulness of behavior.

1. The primacy of purposiveness

The easiest way to settle the issue of primacy would be to find examples of such actions that are intentional but lack one of the two characteristics. Unfortunately, as far as our ordinary practice of ascribing intentionality to human actions goes, it seems that we can find examples for both. It is possible to find examples of intentional, but seemingly purposeless actions as well as actions that are done purposefully, but that are not under the agent's direct conscious control.

[5] In fact, as current research in cognitive science shows, the 'teleological stance' seems to be our 'natural ontological attitude' in the sense that, developmentally, it precedes the attribution of mental states. See the important results of Gergely, Gy. – Csibra, G. 1998, 2003, 2007.

[6] About the importance of the difference between having *the* intention and having *an* intention see Michael Bratman's classic paper, Bratman 1984/1987.

Most human actions in which we are interested are performed for some purpose. But there are some which seem to be intentional but aimless: crossing one's arm in a certain moment or fidgeting with one's pen during a talk does not seem to serve any further purpose. Nonetheless, such actions are still under the agent's direct conscious control, and in that sense, they seem to be intentional.[7]

On the other hand, there are also forms of behavior which are intentional even if the agent has no direct conscious control over them. In fact, the performance of almost all intentional actions has some parts or aspects over which the agent has no direct control. Someone on an airplane may or may not fly intentionally to a certain destination, even if she has no direct control over where the plane will land.

To the latter problem, one can respond that the intentionality of actions requires only that agents can consciously control the initiation of their specific actions – like boarding a specific airplane – even if they lack direct control over every aspect of its performance.[8] However, it is not obvious that an agent must be able to exercise such control even when her behavior is intentional. For, as we shall see, not every form of intentional behavior needs to have a beginning at all. Moreover, not every being which can act intentionally is reasonably assumed to possess the capacity of such control. And further, even when they do, this does not prove the priority of conscious control over the purposefulness of overt behavior in the explanation of the intentionality of actions. For if an agent consciously initiates an action the successful performance of which won't be fully under her direct control, then she must be aware that she is initiating a process that will – if everything goes well – constitute her intentional action. And the relevant process is identified teleologically from the perspective of the end for the sake of which it has been initiated.

Importantly, that an agent *desires* some future event or state that might be the consequence of a process initiated by her cannot explain why what she did was intentional. One can buy a lottery ticket intentionally even if there are many conditions beyond one's direct control which need to be satisfied for one's behavior to count as buying a ticket. But one cannot win a lottery ticket intentionally, even if one desires to win. Buying a lottery ticket is a specific sort of process in which an agent participates with the aim of getting one; while winning a ticket is only one of the consequences of her action, which happens to satisfy her desire.

[7] We might assume that even such actions *can* have some purpose of which the agent is not conscious when performing them. This is certainly possible, but my point is that the actions mentioned *need not* have such purposes in order to understand them as intentional.

[8] Perhaps it is for this reason why Davidson concludes ("perhaps with a shock of surprise") that "We never do more than move our bodies; the rest is up to nature". See Davidson 1980/1971. 59.

For this reason, the teleological understanding of the processes in which agents participate as their initiators is a necessary condition of conscious control over their own behavior. Ultimately, what makes conscious control over one's own behavior possible is that an agent considers what she initiates to do as her action, and hence she must have a prior conception about what counts as an action in a given situation. And what counts as an action in a given situation is determined by the teleological structure of the process envisioned, irrespective of whether or not the agent eventually undertakes the action. Hence, understanding the intentionality of behavior in terms of its purposefulness must be logically and metaphysically prior to the possibility of conscious control.

What then can we say about those actions that seem to be performed without any further purpose? What we need to say is that the mere possibility of such actions does not contradict the idea that the teleological understanding of behavior is a necessary condition of having the *capacity* of conscious control over one's action. For we cannot imagine an agent who performs *all* her actions intentionally only for their own sake. Someone must be able to understand what it means to act in order to do something else or in order to get something before they understand what it means to do something just for its own sake, that is to say, for *no further* end. Hence an agent, who cannot conceive an action, including her own, as a means to satisfy some further end, cannot perform intentionally the types of actions which seem to have no further purpose either.[9]

2. Teleology and the demands of naturalism

In fact, the reason why most contemporary philosophers take the teleological structure of processes that are actions derivative of the exercise of conscious control has little to do with the possibility that actions can be performed for their own sake. For everyone agrees that such cases could not be central for an account of the possibility of intentional behavior. The main reason why so many philosophers consider the purposefulness of behavior derivative of the possibility of conscious control over one's own actions is the conviction that the intentionality of actions must be explicable with reference to their causal origin. And they must be so explicable because there could not be any *sui generis* underived teleology in nature.

If this were right, then the teleological structure of actions must be derivable from agents' prior representations of what is desired or intended to happen and

[9] See Norman Malcolm's important discussion about the very possibilities of intentional 'activities' (Malcolm 1968. 66). Although Anscombe talks about reasons for actions rather than purposes, her famous claim clearly applies in this context as well: "the concept of voluntary or intentional action would not exist, if the question 'Why', with answers that give reasons for acting, did not. Given that it does exist, the cases where the answer is 'For no particular reason', etc. can occur. But their interest is slight" (Anscombe 2000. 34).

the fact that these representations cause what happens when they act. This would imply that even if we typically understand certain forms of behavior as intentional because they have a purpose, they can *have* a purpose only in virtue of the agent's having some prior intention or desire they mean to satisfy. Purposefulness would not be an intrinsic feature of the processes that constitute intentional actions; rather, it would be derivative of agents' having certain types of inner states and the causal role that such states are supposed to play in the production of their behavior.

However, the observation that human actions are often consciously initiated because agents have prior desires and intentions does not establish that the intentionality of actions could be understood with reference to such states. That agents sometimes act *in order to* satisfy their desires does not make the explanation of their behavior nonteleological. Moreover, doing something intentionally in order to satisfy a desire presupposes a prior awareness of the teleological structure of the type of behavior which is performed with that aim. The very thought that an action is executed in order to satisfy some antecedently existing desire presupposes an understanding of the teleological structure of one's own future action.

Further, this approach to the intentionality of behavior entails that only those agents can act intentionally who can also have mental states like intentions and desires; which, in turn, would force us to accept some *a priori* hypotheses about the scope of animal intelligence. Nonhuman animals, just like humans, can act intentionally. If the intentionality of behavior presupposed the possibility of prior mental representations, then nonhuman animals should also be able to represent, and consciously control, their complex intentional actions. Not only monkeys and dogs, but spiders and bees as well. And this seems, for some of us at least, an *ad hoc* and truly incredible hypothesis.[10]

In fact, this *a priori* hypothesis is a direct consequence of the widely shared idea that 'naturalism' in philosophy is compatible only with explanations by prior efficient causes. An explanation of the purposefulness of animal behavior can then be 'naturalistic' only if it hypothesizes that animals have conscious states like (proto-)desires and intentions, which are supposed to be the inner causes of their overt actions. But this idea about 'naturalism' is based on a very limited understanding of what nature is. Teleology can be quite 'natural'; certainly, much more 'natural' than highly contentious *a priori* hypotheses about the mental causes of animal actions are.

In the sequel I shall explain, first, why the Aristotelian account of 'final causes' is well suited for capturing the distinction between intentional and nonintentional forms of behavior. Then I shall argue, negatively, that inner efficient causes cannot explain the difference between intentional and nonintentional

[10] For others, it is not. See especially Steward 2012 (chapter 4).

forms of behavior. From the agent's own perspective, her behavior is intentional only when it is done for the sake of some end; and from the impersonal perspective, an action is intentional if the agent who acts participates in a process with a more or less determinately defined end. An action is unintentional when the teleological process that constitutes an agent's action 'goes astray' in the sense that it fails to reach its end.

II. INTENTIONALITY AND THE NOTION OF ACCIDENTS

Teleological explanations play a fundamental role in Aristotle's philosophy. Modern science emerged as a response to the Aristotelian-Scholastic tradition, and the rejection of Aristotelian teleology became an essential part of that response. This is the origin of the widespread conviction, mentioned earlier, that teleological explanations are incompatible with 'naturalism'. It is, of course, debatable whether contemporary sciences avoid, or should avoid, the use of teleological explanations.[11] My concern here is, however, restricted to the explanation of the intentionality of behavior, not scientific explanation in general.

My thesis is the following: from a broader metaphysical perspective, if an agent φs at *t* intentionally, there must be a sense in which her φ-ing is not a mere accident. It is for this reason that the agent can consciously control what she does in the sense of being the initiator of her own actions. But we can understand the sense in which an action is not an accident if it is intentional only with reference to its final, and not to its efficient, causes. What makes the Aristotelian concept of final causation especially fit for explaining the intentionality of action is not so much Aristotle's own way to apply it in the explanation of natural and social phenomena, but rather his argument for its indispensability: the argument from accidents or coincidences. For it is only the teleological sense of non-accidentality that can explain the difference between intentional and nonintentional forms of behavior.

In one sense, events are not mere accidents if they can be understood as the nomological and/or causal *consequences* of some other events or conditions. However, it is possible that an event is not an accident in that sense but has nonetheless not been performed intentionally. In fact, nonintentional behavior

[11] One of the guiding ideas of the new sciences was that the scientific understanding of the world must be nomic: that the evolution of events is 'governed' by laws of nature. However, there is nothing in the very *concept* of nomic regularity which would entail that laws cannot be teleological. That hearts beat rhythmically in order to help providing the body's cells with oxygen does not seem to be 'less naturalistic' an explanation than it is that blood circulation is caused by regular heart beats. But within the confines of the present paper I am not concerned with the possibility of genuine ('irreducible') teleological laws or regularities. I say more on this in Huoranszki 2022, in Chapters 2 and 4. For the intentionality of a particular piece of behaviour, as I understand it here, does not assume any laws.

is perfectly well explicable nomologically or causally with reference either to the agents' environment or to their internal states. Many types of behavior are explained as mere nonintentional responses to external stimuli or some internal neural changes.

According to the standard 'causal' accounts, agents' actions are intentional if (a) they are the results of some special type of internal causes; and (b) the causal chain leading to the agent's behavior are not 'deviant' (that is, it is of the 'appropriate sort'). In the next two sections I shall argue that (a) the first condition is not necessary for behavior to be intentional; and (b) the second condition cannot be understood without reference to final causes. In this one, I shall explain further in which sense Aristotelian teleology can account for the intentionality of agents' behavior.

1. The significance of Aristotelian final causes

As we shall see, a contemporary teleological account of intentionality need not follow Aristotle's own account of intentional actions in every respect. However, there are at least three important characteristics of Aristotelian teleology or, with the scholastic terminology, of 'final causation' which renders it particularly suitable for explaining the intentionality of actions.[12]

First of all, Aristotelian 'final causes' are not to be confused, as they often seem to be, with *backward efficient causes*.[13] For if backward causation occurs at all, backward causes must actually exist. But the goals or aims for the sake of which an intentional action occurs may never actually come to pass. This means, further, that final causation is not to be understood as a relation between actual events. If it is a relation at all, it is a relation between an agent to whom we ascribe the end and the potential result of a process in which the agent participates in order to reach that end.

Second, Aristotelian final causes are *immanent* in the sense that they are attributed to (animated or inanimate) agents in virtue of their participation in some goal directed processes. Consequently – and contrary, for instance, to the typical Platonic use of teleology – the goal directedness of processes is not explained by some antecedent conscious planning or 'design' which then determines the evolution of events or the shape of human actions. Teleology is not to be understood

[12] The expression "causa finalis" is a legacy of scholastic philosophy. Final causes are one of the four types of Aristotelian causes. However, if we follow Aristotle, it would be more appropriate to distinguish four types of explanatory factors that give different kinds of answers to the question "Why has something happened?". See especially Moravcsik 1974. I use "final cause" because my interest here is not how we can explain what an agent did on a particular occasion, but what makes her behavior intentional.

[13] This important feature of teleological causation is further explained in Hawthorne and Nolan 2006.

as efficient causality in disguise. Aristotelian final causation is a feature of some natural processes that include animal and human behavior.[14]

Third, and relatedly, Aristotelian final causality does not require the truth of panpsychism, neither is it 'anthropomorphic' in the sense that it would involve some illicit projection of human powers and capacities to inanimate substances or to members of some lower species. In Aristotelian final causation the goals or ends for the sake of which an agent behaves in the way she does, need not be antecedently represented by the agent who is involved in the teleological process. And since such aims need not be represented by the agent whose actions they explain, final causation is not to be interpreted as efficient causation by the agent's inner mental states.[15]

This aspect of Aristotelian final causation is crucial for a teleological account of action. As we have seen, no one would deny that intentional behavior must in some sense be teleological: typically, agents act intentionally when their behavior has a goal or aim. But according to the standard version of causal accounts of action, behavior is intentional if it is a causal consequence of the agent's prior or concomitant representation of those aims by her desires or intentions. In contrast, an Aristotelian account of teleology does not require that the aims for the sake of which an action is performed be necessarily antecedently represented by the agent.

I shall mention Aristotle's own example to elucidate the sense of non-accidentality which is, in my view, indispensable for explaining the intentionality of behavior. Suppose a debtor goes to the market in order to buy some goods. The creditor, who has long desired to get her money back, goes to the market to sell tickets to a feat. As it happens, they meet, and the debtor pays back then and there what she owes to the creditor. Thus, the desired or wanted consequence occurs. But it occurs as an accident in the sense that the creditor did not go to the market for the sake of getting her money back (Aristotle, *Physics* ii 4, 196b).[16]

Importantly, the same would be true, if the agents did not go to the market intentionally but were taken there by some brute force. Even if the creditor has some desire and hence some end in view when he goes to the market, the event of encounter necessary for reimbursement did not happen *for the sake* of satisfying that desire.

[14] About the Platonic understanding of teleology see Lennox 1985. About natural theology as a form of denying immanent teleology, see Johnson 2005. 30–35.

[15] For a critical overview of the attempts to reduce Aristotle's final causes to efficient ones see Charles 2012. 235–238.

[16] For further relevant examples and an alternative interpretation of the problem see Sorabji 1980. 3–26. According to Sorabji's account, accidents or coincidences have no causes. But this seems plausible only if "causes" are restricted to Aristotelian final causes since the event of encounter has obviously some efficient (not to mention some material) cause.

My suggestion is that this Aristotelian example of an accident captures the essence of the sense in which behavior can be intentional. Davidson's question, with which I began, clearly indicates that the intentionality of actions can only be understood in contrast to those episodes in an agent's life that *merely happen* to her. But what does it mean that things 'merely happen' to an agent? Such 'mere happenings' can certainly have prior 'efficient causes'. What 'merely happen' to an agent are those episodes in her life which are accidents or mere coincidences in the sense that, although they have prior causes which can explain why they have happened, they do not happen for the sake of an end.

2. Normativity and mental causes

Thus, as I see it, Aristotelian teleology can capture well the sense of non-accidentality that is the essential feature of the intentionality of behavior. This does not mean, however, that I propose to follow Aristotle's own account of intentional actions in every respect. There are at least two aspects of my proposal in which it diverges from Aristotle's own account.

First, the ascription of final causes in the sense I shall understand them in the present context does not have any *direct* normative implication. The fact that an action is performed for the sake of an end does not in any way justify what an agent does. Put otherwise, that an agent's behavior has final causes does not imply that it was good for the agent to act in that way.

Final causation need not justify an agent's behavior even in the weak sense that what she has done must always be interpreted as a *good* means to achieve an end. When I lose my sense of direction, I may start walking towards the north, even if my end is to reach a place that lies south of where I am. My behavior was then intentional rather than nonintentional, even if it is a most inefficient means to achieve the end for the sake of which it occurred; and even if I unintentionally ended up in a place I did not intend to.

Aristotle himself often attributes goals on the ground that they are good for the agent in the sense that the agent benefits from the satisfaction of the ends for the sake of which she acts. But the application of Aristotelian teleology, particularly in the context of intentional actions, does not require the use of this metaphysically more loaded notion of final causation. The ascription of final causes itself need not have such normative implications. When behavior occurs for the sake of an end this is a *fact* about it; irrespective of whether or not the end is reached or whether or not it is reached by some more or less effective means; and irrespective of whether or not the agent should have that end at all.[17]

[17] The question about normativity is further complicated by the fact that Aristotle distinguishes between real and apparent good, and one might want to argue that every goal must

Nonetheless, Aristotle is certainly right to the extent that it is a necessary precondition of the rational and moral evaluability of actions that they are done for the sake of some end. An agent's action is rational if she chooses it as the most efficient means to satisfy the ends for the sake of which she acts; and an agent's action is good if she follows the end(s) for the sake of which she ought to act. The teleological structure of actions can provide the basis of their normative evaluation.

Second, Aristotle's own way of applying teleology in the explanation of intentional behavior seems to be mentalistic. Aristotle himself holds that animals who lack rational capacities can still act intentionally because they have some 'thought and desire' which direct their behavior towards some ends.[18] As mentioned earlier, whether or not we want to follow him in this respect depends on whether or not we find plausible the idea that all animals that are capable of intentional actions – not only cats or dogs, but also ants, flies or bees – have 'thoughts and desires'. I find such mental ascriptions entirely *ad hoc* in most cases. And an Aristotelian-teleological understanding of intentional action does not require it.

It is for this reason that I prefer using "for the sake of which" to express the goal-directedness of an agent's behavior, even if the use of "with the intention that" may sound more natural. The use of "with the intention that" intimates that the intentionality of behavior requires some prior or concomitant intention with which the action is performed; or that the action must have been *intended* by the agent. In fact, even most contemporary non-causal accounts of action assume this.[19] They agree with the efficient-causal accounts in that the explanation of the intentionality of actions must be mentalistic.

However, one of the central aims of this paper is to argue that an account of the goal-directedness of behavior need not be mentalistic. The explanation of how goal-directed behavior can be consciously controlled by the agent partici-

---

be understood at least as apparently good. However, I have two concerns about applying this distinction in the present context. The first is that when I put salt instead of sugar into my coffee in order to sweeten it, I do something intentionally which does not at all appear to me good. It might be true that salt appeared to me to be sugar, but this does not mean that it appeared to me good to put salt into my coffee. The second is that the very concept of apparent good presupposes a mentalistic understanding of teleology which, for reasons I shall present in the next few paragraphs, I reject.

[18] "Now we see that the living creature is moved by the intellect, imagination, purpose, wish and appetite. And all these are reducible to thought and desire" (Aristotle, *Movement of Animals*, 700^{b}15). For a contemporary account of intentional action similar in this respect to Aristotle's see Hyman 2015. 106–111. Importantly, for Aristotle, "thought" need not entail the use of the intellect, which is a rational capacity, and which is characteristic only of humans. But it does presuppose the capacity of memory, perception, and desire which is the inner mental cause of action.

[19] See for instance von Wright 1971, Wilson 1989, or Ginet 1990. A rare example for an early non-mentalistic teleological account of action see Collins 1984.

pating in it is, of course, mentalistic – what else could it be? But the question about the possibility of conscious control should not be confused with the question concerning what makes behavior intentional in the first place.

Of course, we can often explain *why* an agent did something by ascribing some intention or desire to her; and then we assume that the agent who performed the action must have had some end *in view*. My point is that the ascription of such states does not explain why what the agent did was intentional. What explains the intentionality of actions is always the fact that the agent's behavior occurs for the sake of some end. In certain cases, those ends need not even be represented by the agent; whereas in others, even if they are represented, this does not explain why the behavior that may satisfy them is intentional.

This is not to deny the importance of agents' intentions in the explanation of their behavior. Rational agents can exercise conscious control over their own behavior only if they are able to choose action with some end in view. To have an aim in mind, together with a plan about how to achieve that aim, is necessary for the exercise of conscious control over one's own behavior.[20] But, again, even if the capacity to represent one's own aims is necessary for rational guidance, what explains the guided behavior's intentionality is not its efficient, but its final cause: the fact that the behavior did occur for the sake of an end.

III. THE INDISPENSABILITY OF TELEOLOGY

In this section then, I shall argue that, as far as the explanation of the intentionality of behavior is concerned, final causes are indispensable and irreducible to inner efficient causes. First, as the possibility of animal agency shows, we can ascribe ends to agents and to the processes in which they participate without assuming that those ends are antecedently represented by the agents who act. But more importantly, even when the relevant aims can indeed be so represented, we cannot understand the intentionality of behavior without reference to the intrinsic teleological nature of the processes in which the agent who acts participates.

1. Actions, movements, and the agent as a cause

As Aristotle already observed, animals are self-movers. Following this observation, some recent accounts of action suggest that we can understand the intentionality of behavior with reference to the exercise of agent's capacity to initiate

[20] For questions about how representations of aims can causally guide actions see, among others, Bratman 1987 and Mele 1992.

their own movements.[21] Such accounts note that there is an interesting ambiguity in the use of some English verbs which can describe an agent's action.[22] "Move", for instance, can be used both transitively and intransitively. When we say that an agent's arm moves, we describe an event which may or may not be the agent's action. It seems, however, that saying that *the agent moves her arm* entails that what she does is her action. And it is her action because the movement of her arm is the *causal result* of what she does. This observation about the use of some action-verbs seems to countenance the idea that bodily movements are intentional if, and because, they are caused by the agent as a mover.

There is another observation about the language of action which can be invoked in support of the idea that agency is manifested on those occasions when an agent is causing her own movements. Sentences which have an agent as their grammatical subject and contain a transitive verb need not describe intentional actions. They can, for instance, describe perceptual states or processes. However, the use of verbs describing perceptual states and processes does not reflect *the direction of efficient causality* that may be involved in the processes of the acquisition of such states. An agent can see, hear, feel etc. certain things; but she does not thereby cause them to be seen, heard, or felt. In contrast, the use of verbs expressing intentional movements does seem to reflect the direction of efficient causation. When an agent *raises* her hand, she causes the movement of her body.[23]

It can be objected that such observations about the language of action cannot be generalized, since many verbs – like running, crying, flying, or writing – can unambiguously describe actions even when they are not used transitively.[24] But one can retort that the performance of any such action must involve, in one way or another, some bodily movements. Since overt physical behavior cannot be performed without the movement of the agent's body, one can insist that behavior is intentional if and only if the relevant movements have been caused by the agent as a mover. Bodily movements themselves, when "movement" is understood intransitively, are events which are the *results* of the agent's moving her body. Such movements manifest agency only to the extent that they are parts of the agent's causal activity: her moving the body.

If this line of reasoning is correct, then we cannot understand physical actions without relying on the prior concept of the agent as a mover or a causer. Behavior

[21] This observation plays a crucial role in Helen Steward's account of agency, see Steward 2012. 71–72.

[22] See Hornsby 1980.

[23] This is also the ground of trying to categorize certain mental states with reference to their 'direction of fit'. See particularly Searle 1983. My point here, of course, is *not* about mental states, for the theories I discuss emphasize the role of the *agent* in the etiology of intentional action in contrast to the agent's states.

[24] As it has been noted in an early review of Hornsby's book by Watson 1982.

is intentional in virtue of having a peculiar kind of efficient cause. Since physical movements are results and hence effects, they must have prior efficient causes. The metaphysical problem of agency and action seems then to boil down to the question about the nature and operation of such efficient causes; that is, to an account of how agents as persisting substances can cause their own behavior.

It seems certainly right that – as our language of action suggests – agents are typically self-movers. And I see no good reason to deny, as some may do, that agents can be the causes of at least some results of their own actions. Moreover, on this view, just as in the teleological account, actions are not taken to be momentary *events* like instantaneous movements of the body, but *processes* with some results.

However, even if it is true that when agents act, they often exert their causal capacities and thereby cause certain events to happen, this cannot explain the difference between intentional and nonintentional forms of behavior. For even if agents are typically self-movers, an agent's causing the movement of her own body is neither necessary, nor sufficient for the intentionality of her behavior.

2. Agent-causation and the intentionality of actions

The first thing to note about the capacity of self-movement as the explanation of intentionality is that not only agents' actions, but also their omissions can be intentional (as opposed to being *non*intentional). Obviously, intentional omissions do not involve the agent as a mover at all. Thus agent-involving efficient causation cannot be *necessary* for the intentionality of behavior in general. According to the testimony of *Phaidon*, Socrates stayed intentionally in Athens after his condemnation, even if, as he says, he could have already been in Megara (Plato, *Phaidon*, 99a). And that he remained in Athens in the circumstances in which he did certainly manifested his agency because it was a form of conscious intentional behavior.

In many moments in their life, agents do things intentionally without moving their own body. And even more often, whether and how they move their body is simply irrelevant to the intentionality of their behavior.[25] If Socrates had chosen to escape, his action of escaping would have been intentional even in the moments of motionlessly sitting on a cart on his way to Megara. In general, people can do intentionally many things without thereby causing anything to happen. Intentional agency can be manifested even in those moments when agents do

[25] Attempts have been made – for instance by Smith 2010 – to answer this problem by saying that whenever the agent intentionally omits to do something then she does something else. But that is entirely irrelevant, since whatever we do, we omit countless other things unintentionally. See also the exchange between Carolina Sartorio and Randolph Clarke in Aguilar and Buckareff 2010.

not move their body; or rather, when the movements, even if they are caused by the agent, are simply irrelevant to what they do intentionally.

But further, and more importantly, even on those occasions when the performance of an intentional action does require that an agent be the mover of his own body, his being the efficient cause of his own movements cannot explain why the movement was intentional. Many persisting substances that are incapable of acting intentionally can still be the causes of their movement or some changes in their surroundings. The hemlock poisoned Socrates thereby causing his death; sugar sweetened my coffee (caused it to become sweeter by dissolving in it); and my alarm clock wakes me up by making (causing) that terrible noise in the morning. In fact, it is arguable that the causative use of transitive verbs is the most common way to express causal claims.[26] And when we express a causal claim in this way, we assign a causal role to a substance. It seems then that substances can be causes even if they are not able to act intentionally.

Human agents are, among other things, persistent physical and biological substances with many causal powers; and they can, merely in virtue of being such substances, cause many kinds of events. I can break a glass, make a noise, stir the air around me, and warm up a bed without acting intentionally. Some of the things that I cause, I cannot do intentionally; others I can, but I might cause them only accidentally. And the same is true even in those cases when the action's results are the movements of my own body.[27]

Here is an often-discussed case. Suppose a neurologist taps my knee with her rubber mallet. Then, as a spontaneous neural reaction, I move my leg. My moving of the leg manifests my power to move it in certain circumstances; and further, it bears witness of my – in this respect at least – properly functioning neural system. But the movement was not intentional even if *I* did raise my leg and even if my leg's movement was a result of my moving it.

Advocates of the agent-causalist account of intentional action may reply that, whenever an agent's movement is 'only' a reflex-response to a stimulus, 'merely neural and muscular processes operate'. But even if this is certainly right in a sense, that can hardly explain why my behavior was not intentional. Presumably, whenever an agent performs an overt physical action, intentionally or not, neural and muscular processes operate. Saying that nonintentional movements are the results of some *merely* neural and muscular processes cannot explain the difference between them and intentional movements since the question is pre-

[26] See Anscombe 1993, Strawson 1985, and Lowe 2009.

[27] According to some versions of the agent-causal account of actions – like, for instance, Taylor 1966 or Clarke 2003 – agents cause their own actions, not the movement of their body (which is the result of the action). However, for reasons well exposed by Hyman and Alvarez 2002, Hornsby 2004, and Lowe 2009 those versions of the agent-causal view do not seem to be coherent.

cisely why bodily movements caused by the agent are 'merely such and such' in certain cases while manifest intentional agency in others.

Perhaps one would want to deny that in the case described *I* moved my leg, because although my leg indeed moved, it was not *me* but the neurologist who moved it (by tapping my knee with her rubber mallet). I was, as it were, a *mere patient* in this process. However, neither the emphasis on personal pronouns nor a more detailed inspection of the causal *history* of my movement can answer the problem here. For even if the neurologist's action was, in the circumstances, *a* necessary causal condition of the movement, it would be bizarre to claim that thereby *my* raising the leg was *her* intentional action. What the neurologist wants to check by tapping my knee is whether or not, when my knee is hit, *I* shall move my leg. She is not interested in whether or not *she* can move it (by being able to strike a strong enough blow on it, for instance).

Thus, the transitive and causative use of the verb describing a patient's behavior is as essential here as it is supposed to be in the case of intentional actions. Even if the movement of the leg was a causal consequence of what the neurologist did, this does not show that the patient has failed to be the mover of his own body. His causal contribution was as necessary for the movement in this case as it is when the doctor *asks* him to raise his leg in order to check whether or not he can do so. In both cases, the doctor's action might be a causal antecedent of the movement of the patient's leg. But in neither case would the movement have occurred without *the patient* moving his leg. In fact, even if the neurologist indeed caused the movement of the patient's leg, it would be wrong to say that *she moved his leg* instead of him.

Imagine, further, that before the patient goes to the doctor, he is aware of the purpose of the test, and that he wishes or desires that he raise his leg as a response to his knee being hit. In this case, he had a desire that has been satisfied by his own causal activity. But all this does not make his behavior intentional. And the reason why his behavior was not intentional is that he did not move his leg *for the sake of* that end, never mind how much he wished or desired that the movement occur. In fact, curiously, if he had moved his leg for the sake of that end, his wish or desire could not have been satisfied.

In sum, an agent can be a self-mover and hence the cause of her own action without acting intentionally. And conversely, an agent's behavior – like Socrates' staying in Athens – can be intentional and manifest agency without the agent causing anything. Whether or not behavior is intentional depends on whether or not it was performed *for the sake of* an end that we can ascribe to the agent and hence to the process in which he participates. Moving one's own body and hence being in this sense the efficient cause of one's own behavior is neither necessary nor sufficient for manifesting intentional agency.

3. Volitions and intentional actions

Our considerations in the previous section can be summarized like this. We observe that many verbs expressing overt physical actions are transitive and causative. This suggests that agents can be considered as efficient causes of their actions' results. What we have seen is, however, that mere reference to the agent as a cause cannot explain the difference between intentional and nonintentional forms of behavior.

We may seek to remedy the weakness of the purely agent-causal account by specifying some kind of internal event that the agent can cause directly, and the causal operation of which can guarantee the intentionality of overt behavior. The problem with the purely agent causal account of intentionality might be that it takes the agent to be the direct cause of the movement of her body. But agents as persisting substances can cause the movement of their body in many different ways.

We might want then to specify a pertinent way in which an agent must cause her own behavior in order to make it intentional. We might say that whenever an agent acts intentionally, she causes *directly* some sort of event which occurs 'inside' her and by which she initiates the movement of her body and hence her overt actions. The agent's overt behavior is intentional if it is a causal product of the occurrence of that sort of event. Otherwise, the behavior is nonintentional.

As we have seen, (efficient-)causal accounts of intentionality are grounded in the assumption that whenever agents act intentionally, they exercise some control over what they do. But for the exercise of the pertinent kind of control it is not sufficient that they as persisting substances cause their own behavior. They must cause it through exercising direct control over the occurrence of a special kind of internal event. Agents cause changes in their environment by moving their body. But they do not cause the movement of their body directly. Rather, they cause them by causing first the occurrence of an internal event, which is then the event-cause of their external behavior.

Then, the intentionality of physical behavior might be explained by the fact that intentional movements have been caused indirectly by the agent's first causing something else directly. Reflex behavior is not intentional because it is not caused by the agent's causing first a kind of event which is the necessary causal antecedent of every movement that is intentional. Although there are other ways to identify the relevant sort of event that the agent might directly cause, it shall serve my purposes here to follow a long tradition and call such events as 'conscious volitions' or 'acts of will'.[28]

[28] For a useful summary of the modern history of volitional theories see Hyman 2015. Chisholm claims that the agent can directly cause a cerebral event; others (like O'Connor

There is a standard objection to the volitionist accounts of intentional action which I set aside for the moment. Ryle has famously argued that if we understand volitions themselves as actions, then the volitionist account leads to a vicious infinite regress; whereas if volitions are understood as episodes that merely happen to the agent, then they cannot explain the intentionality of behavior.

Yet, even if this is indeed an objection to the idea that *every* event is an action in virtue of its causal origin, it does not show why the intentionality of overt behavior cannot be explained by its volitional origin. Perhaps volitions are intrinsically actions; or perhaps they are actions in virtue of being directly caused by an agent. In either way, overt behavior might be intentional because by willing an agent causes the results of her volition.[29]

Nonetheless, a fundamental problem remains. Any event, and particularly things that agents do, can have many actual consequences. It is hard to see why volitions as internal psychological or psychophysical events would be different in this respect. It is obvious though that not every causal consequence of such events is an action. Willing to perform an action can result in many psychological, physiological and behavioral changes (excitement, rising blood pressure, trembling hands) which are not intentional. On the volitionist account, behavior is intentional if it is part of a process initiated by the agent's volitions. But since there are probably always many sequences of events that are initiated by a psychological or psychophysical act of volition, we need to explain why only certain causal consequences of this volitional act or event are intentional.

One may think that the explanation is very simple: the willed physical behavior is intentional only if it is 'content matching' in the sense that only those consequences of willing should count as the agent's actions that somehow 'match' the content of her volitions. However, and crucially, the content of a volition cannot be the movement of the body (or some other result of the action). It seems obvious that one can will – as opposed to wish, desire, or hope – to perform only actions that are intentional. Thus, volitions cannot explain the intentionality of the acts willed by the agent. They presuppose it.

Imagine that someone desires or intends to replace a table. This does not mean that she can will that the table move from one place to another. A person who could achieve that a table moves simply by willing that *it* moves would do simple magic. What an agent can will is to move the table; that is, to initiate, and participate in, a process with the end of the table's being replaced. In general,

2000) talk about the agent's causing action-triggering intentions. These accounts differ from each other in detail, but these differences are largely irrelevant for my point here.

[29] About the standard objection see Ryle 1949. 62–75. For the different versions of volitionist accounts of intentionality see McCann 1974, McGuinn 1982, Ginet 1990, and Lowe 2000.

the agent can only will to perform an intentional action with a certain result. Similarly, the agent cannot will that a hand of her rise; she can only will to raise a hand; that is, to do something (mentally and then physically) for the sake of her hand's rising.

Hence, although it might be true that reflex responses are not willed by the agent, this does not explain how volitions can make actions intentional. It is rather the other way around: one *cannot will to* perform a reflex response because a reflex response is a kind of nonintentional behavior. Similarly, one cannot will to perform accidentally a bodily movement or any other action, because neither accidental movements nor their consequences can be brought about intentionally.

This means that if volitions are psychological events with content, then their content can only be intentional actions; and hence they can hardly explain the very intentionality of actions. If acts of volitions occur at all, they occur because they are the initial parts of some behavior which is performed for the sake of some end. Acts of will cannot explain the intentionality of an agent's behavior; rather they too are explained by the end(s) for the sake of which they occur.

Consequently, while the possibility to will an act may help explain how an agent can consciously control what she does, it cannot account for the very intentionality of the action done. The possibility of volitions as internal mental events presupposes, rather than grounds, the intentionality of certain forms of behavior that the agent might be able to control by willing to perform it. The ground of intentionality still seems to be that the agent's behavior, which may or may not be subject to her conscious control, has been performed for the sake of an end.

IV. TRYING AND THE MODALITY OF INTENTIONAL ACTIONS

So far, I have argued that the intentionality of actions can be explained only teleologically. A piece of behavior is intentional only if we can identify an end for the sake of which it is done. Similarly, conscious omissions are intentional in the same sense: we can consider an agent's omission as a form of intentional behavior only if the agent omits an action that she would otherwise be able to perform in the circumstances for the sake of achieving some end.

However, it might seem that even if an action cannot be intentional unless it is done for the sake of some end, the teleological structure of a process in which the agent participates cannot be sufficient for explaining the difference between intentional agency from nonintentional one. My heart plumps blood for the sake of providing the cells in my body with oxygen. Nonetheless, my heart cannot act intentionally; only I as an agent can. Moreover, although I am an agent, not

everything that I do for the sake of an end is an exercise of my agency. When I run for a while on a hot day, I start sweating. I sweat in order to cool down my heated body. Nonetheless, the secretion of sweat is not my intentional action.

In fact, it is partly such examples which may lend support to the view that it is at least necessary for a kind of behavior to be intentional that it be causally initiated by some of the agent's inner mental states. However, in this last section I shall argue that we need not turn to the causal-mentalistic hypothesis to explain the distinction between teleological processes which are the agents' intentional actions and teleological processes which merely involve an agent without being her actions.

Davidson once argued that behavior manifests agency "if what [the agent] does can be described under an *aspect* that makes it intentional" (Davidson 1980. 46, my emphasis). I suggest that there is a special aspect or mode of the teleological processes in which an agent participates that explains why they are also the agent's actions. Whenever agents φ intentionally, it must be true that they also try to φ. Trying to φ seems to be the universal aspect or mode in terms of which actions can be redescribed if they are done intentionally. Similarly, "trying to…" is also the special mode or aspect of omissions that explains how they can be intentional.

I need to address two objections to this idea. According to the one, we cannot say of every intentional action that the agent who performs it also tries to do it. In fact, if the objection is correct, we can say of an agent that she is trying to do something only in special circumstances. For "trying to" applicable only when an intentional actions have failed to reach their aim.

According to another objection, "trying" is merely an interpretation of "willing" and hence trying is not a special mode of description that the intentionality of action entails, but rather the initial phase or the mental antecedent of the intentionally performed bodily movements. If this were right, then my claim that the possibility of intentional actions is logically/metaphysically prior to the possibility of conscious initiation of such actions would be wrong. Since then, bodily behavior would be made intentional after all by its necessary mental-causal antecedent, and not by the intrinsic telic feature of the processes that constitute an agent's intentional action.

We can raise this second objection in another way as well. Suppose that the first objection is answerable and hence whenever an agent does something intentionally, she also tries to do it. What explains this? One possible explanation seems to be that an overt action can be intentional only if it has an initial mental phase which consists in the agent's merely trying to perform the overt action. So interpreted, trying to φ is an action that is performed 'within the agent's skin' or 'within the spatial envelop of her body' before her body begins to move. Hence trying to φ would always refer to some psychological or psychophysical action that precedes the overt physical behavior. Trying would not be the aspect or

mode under which every overt intentional action can be described. Rather, as in the mentalistic accounts, actions would be intentional because they begin with the agent's mentally or psychophysically trying to perform them.[30]

1. Two senses of trying

Now, I am not disputing that there are special cases in which trying *can* be understood as an agent's purely internal and/or mental action. But it can be so understood only if it is conceived as an initial phase of a more complex process that *would* constitute an agent's action if the circumstances were 'normal'. It does indeed seem plausible that an agent could have done *something* intentionally even in those cases in which her intended action was aborted at its initial phase when no overt physical movement has yet occurred. And it seems that whatever the agent did in such cases she could have done it only internally and perhaps mentally. We may want to say then that what she did was 'mentally trying' to perform an action that, in normal circumstances, she would have performed physically.[31]

However, from the fact that sometimes agents can try to do something even when they do not perform any overt physical action, it does not follow that we need to understand trying in this way in every case. An agent can also consciously omit to do certain things (for instance, join the army) and thereby she can try to do (achieve) certain things (for instance, to stop a war), but this does not mean that she 'merely' tries to do so in the sense that she would be unable to consciously control how she acts physically.

Earlier in the first section, I argued that we must distinguish two different senses in which an agent's actions can be said to be 'intentional'. In one sense, the intentionality of an action is contrasted with what is nonintentional; in another, it is contrasted with what is unintentional. An action is intentional in the first, more fundamental, sense if the process that constitutes it has a certain teleological structure so that it is performed by the agent for the sake of an end; while an action is intentional in the second sense when it achieves what the agent has in mind by initiating it.

Similarly, and relatedly, we also need to distinguish two senses of trying to do or trying to get something. In one sense, trying can indeed be understood as a kind of mental action: it is the initial 'inner' – that is to say, not, or not yet,

[30] See Armstrong 1968 and O'Shaughnessy 1973. In Hornsby 1980 we can find a similar account of trying. Although McGinn 1980 and Ginet 1990 talk about willing rather than trying, their views admittedly have certain affinities to the idea that bodily movements are the results of the agent's trying to act. Searle's analysis of intention in action in Searle 1983 has also been interpreted as a version of the trying-theory by Mele 1992, and later by Searle himself in Searle 2001. See also Lowe 2000. 246–252. For a meticulous criticism of such accounts, see Cleveland 1997.

[31] This argument originates in William James' famous case about the patient with anesthetized hand in James 1890/1983. 1101–1102.

overtly physical – part of an intentional action. It is in this sense that trying to φ when an agent φs intentionally is also a condition of the possibility of agents' conscious control over their own behavior.

In another sense, however, trying is not meant to refer to the initial, merely inner phase of an action. It seems a perfectly good answer to the question "Why do you push that button on your keyboard?" to say that "I try to save my document". My trying to save the document then consists in an overt action of mine (pushing the button) and not in an inner mental act. My intended action is complete only when my document is saved, but by moving my finger in the way I did I also tried to do what I could in the circumstances in order to save, or for the sake of saving, my document.

In a more fundamental sense then, trying means that an agent does everything she can in circumstances C in order to φ or for the sake of φ-ing. In this sense, trying is the special aspect or mode of describing intentional actions because it captures their specific teleological structure and hence the aspect under which they are intentional. But do indeed all actions that are intentional entail that the agent tries to do them? Is trying 'ubiquitous'?[32]

2. The ubiquity of trying and the teleological structure of actions

Observations about how we normally talk about actions do not seem to support ubiquity. For although we often say that agents tried to do something which they have eventually failed to do, we rarely say that they tried to do what they have succeeded in doing. The rare exceptions are when agents must overcome some challenge or when, for some reason, the initial likelihood of failure is relatively high.[33] But such cases aside, that is, in all cases when success is not surprising and the action has been accomplished, it sounds strange to say that an agent tried to do what she has done.

So, we need an explanation of why it is true that an agent tries to φ whenever she φs intentionally. And we also need an explanation of why what we tend to say is not decisive in this matter.[34] One possibility is, again, to return to the idea that every overt physical action that an agent does intentionally must be

[32] The idea that trying is ubiquitous were introduced by Hornsby 2010. Hornsby says more recently that "even if trying to φ is a necessary condition of intentionally φ-ing, still trying to φ does not introduce any causal element into intentionally φ-ing" (Hornsby 2010. 22). Rather, she claims that "to *try* is to *do what one can*" (Hornsby 2010. 20). See also Cleveland 1997 and McLaughlin 2012. 114.

[33] The original point is made by Wittgenstein, see *Philosophical Investigations* 622.

[34] There are many truths we would not mention explicitly because, mentioning them would have inappropriate implications in a given context. This argument, which relies on Grice's account of 'conversational implicature', has been first applied to trying by O'Shaughnessy 1973. The argument has been challenged by Watson 1982. For a recent defense see, again, Hornsby 2010.

preceded by her mental action of trying to do it. But this idea is mistaken. For the intentionality of actions *itself* does not entail anything about an agents' antecedent mental activity.

First of all, φ-ing intentionally entails that agents also try to φ even if there is no reason to assume that they can consciously represent their intentional actions before they perform it. Animals and toddlers can try to do things even if the inner phase of their actions cannot be described as 'mental trying'. A spider can try to spin a net in my study even if my cleaning activity aborts the attempt. A toddler can try to walk to her mum, even if she does not yet have any conscious representation of a process that *we* can describe as "walking to her mum".

Moreover, the intentionality of many forms of adult behavior cannot be understood with reference to a mental action that is the initial phase of the agent's physical movement. Someone on an airplane can try to reach a certain destination and hence can fly there intentionally even in those moments when she sits motionless on the plane; or even if the plane eventually lands somewhere else. Trying to do something does not require the exercise, or even the possibility, of active conscious control. Neither does it seem to require a prior mental representation of one's own behavior as a future action that the agent tries and hence starts to perform.

Consider one of Davidson's often cited examples when, on a particular occasion, someone moves his finger, flips the switch, turns on the light, illuminates the room, and alerts a burglar (Davidson 1980. 4). The descriptions of such actions are related in the following manner: the agent illuminates the room *by* turning on the light, turns on the light *by* switching the flip, switches the flip *by* moving his finger, and so on.[35] If we understand these action-descriptions as referring to parts of a teleological process (rather than to an instantaneous event as in Davidson), then we can also express their connection from the 'opposite direction' as it were: the agent flips the switch *in order to* turn on the light, turns on the light *in order to* illuminate the room, and so on.

However, given the ubiquity of trying, the agent in this example must also try to move his finger, try to flip the switch, and try to turn on the light. Suppose now that 'trying to φ' is the initial mental phase of the action. Must then there be three (or more) antecedent mental actions that precede the movement of the

[35] Davidson used this example in order to support his claim that he has performed only one action which can be individuated in different ways in terms of different results. In this he follows Anscombe 2000, who is also followed by Hornsby 1980. Others (for instance Goodman 1970, Ginet 1990, Alvarez and Hyman 1998) would say that the different descriptions in the example refer to different actions. But all these accounts assume that actions are either events or the causings of some events which are their results. According to the teleological account, however, actions are processes so that the descriptions in the example refer to different phases of the same process. About actions as processes see especially Thompson 2008.

body? Or three initial mental phases of the process that constitutes his action?[36] But then, how are those 'mental tryings' are related to each other and the subsequent overt action?

It seems right that the agent moved his hand intentionally because he tried to flip the switch in order to turn on the light and in order to illuminate the room. But it is hard to make sense of the view that thereby he performed the mental (or psychophysical) action of trying to move his finger in order to perform the mental action of trying to flip the switch and the further mental action to try to illuminate the room. If trying is understood as a mental action, then trying all those things must be one and the same action. But then, what the agent is trying to do initially or mentally is the whole process of his intentional action with a given teleological structure. The content of his 'mental trying' must be the whole action at once.

Consequently, trying to φ is to be understood as a mental antecedent of overt intentional behavior only in exceptional cases. "Trying to φ" *can* express the initial phase of the process of an intentional action which the agent does only mentally (or psychophysically), but it need not. And hence trying is not ubiquitous because an action can be intentional only if it has an initial mental phase. Rather, whenever an overt intentional action occurs, what an agent tries to do is something that she does physically, something that has been accomplished in way of doing what she aims to do or achieve. Trying to turn on the light *is* for the agent, in the given circumstances, to flip the switch because she flipped the switch for the sake of turning on the light, irrespective of whether or not the light has eventually been turned on.

Trying is ubiquitous because intentional actions are processes with a special teleological structure. An agent tries to illuminate the room by flipping the switch in circumstances in which flipping the switch is necessary for illuminating the room; and he tries to flip the switch by moving his finger for the same reason. When an agent has tried to φ she did everything she could in the circumstances in order to f. It is in this sense that "trying to φ" expresses the mode or aspect of actions that make them intentional.[37]

Interpreting trying in this way explains not only its ubiquity, but also why we mention it so rarely that an agent tried to do what she did successfully. Suppose Socrates sits on his bed while he talks to his friends. It follows from this that he *can* sit on a bed while he talks. But we would mention that he can only in special

[36] This problem, let me emphasize, arises no matter how we answer the question about how many actions such descriptions describe. The issue is not how we individuate actions, but how logically/conceptually "doing f intentionally" and "trying to φ" are related.

[37] Hornsby says more recently that "even if trying to φ is a necessary condition of intentionally f-ing, still trying to f does not introduce any causal element into intentionally φ-ing" (Hornsby 2010. 22). She also claims that "to *try* is to *do what one can*" (Hornsby 2010. 20). See also Cleveland 1997 and McLaughlin 2012. 114.

circumstances; for instance, when he lies in his bed and we wonder why he does so; or when, for some reason, we are astonished that he can sit. Nonetheless, since actuality entails possibility, if he does sit on his bed while he talks, it is certainly true that he *can* do so.

Similarly, if Socrates sits intentionally where he does, then he also tries to sit there, even if in most circumstances it would sound weird to mention this. But this does not entail that trying to sit on that place is a mental action by which Socrates performs his sitting on his bed; neither is it a strange way to describe what he does. It is the modal consequence of his sitting there intentionally, which we mention only in specific contexts.

In a sense then, trying is ubiquitous because the truth that an agent tries to φ is a modal consequence of her φ-ing intentionally. And trying to φ is a modal consequence of φ-ing intentionally precisely because saying that an agent tries to φ is a way to specify a goal or aim for the sake of which the agent's behavior occurs. When an agent acts intentionally, she tries to do something with some result. Trying to φ does not imply φ-ing, because trying to φ specifies the end for the sake of which a kind of behavior is performed in given circumstances irrespective of whether or not the action has been accomplished, and hence irrespective of whether or not that end has ever been reached.

However, φ-ing intentionally does imply trying to f in the sense in which "trying to" is the most general way to identify an agent's ends at performing some actions with reference to some particular result for the sake of which she behaves in the way she does. It is for this reason, and it is in this sense, that trying is *a*, perhaps *the*, mark of exercising agency as it is manifested by intentional behavior.

REFERENCES

Aguilar, Jesus H. – Andrei A. Buckareff (Eds.) 2010. *Causing Human Actions*. Cambridge/MA, The MIT Press. https://doi.org/10.7551/mitpress/9780262014564.001.0001

Alvarez, Maria – John Hyman. 1998. Agents and Their Actions. *Philosophy*. 73. 219–245. https://doi.org/10.1017/s0031819198000199

Anscombe, Elizabeth 2000. *Intention* (2nd ed.). Cambridge/MA, Harvard University Press.

Aristotle 1984a. Physics. In Jonathan Barnes (Ed.) *The Complete Works of Aristotle*. Princeton, Princeton University Press. 315–446.

Aristotle 1984b. Movement of Animals. In Jonathan Barnes (Ed.) *The Complete Works of Aristotle*. Princeton, Princeton University Press. 1087–1096.

Armstrong, David 1968. *The Material Mind*. London, Routledge and Kegan Paul.

Bishop, John 1989. *Natural Agency*. Cambridge, Cambridge University Press.

Bratman, Michael 1984/1987. Two Faces of Intention. In Bratman 1987. 111–127.

Bratman, Michael 1987. *Intention, Plans, and Practical Reason*. Cambridge/MA, Harvard University Press.

Charles, David 2012. Teleological Causation. In Christopher Shields (Ed.) *The Oxford Handbook of Aristotle*. Oxford, Oxford University Press. 227–266. https://doi.org/10.1093/oxfordhb/9780195187489.013.0010

Clarke, Randolph 2003. *Libertarian Accounts of Free Will*. New York, Oxford University Press. https://doi.org/10.1093/019515987x.001.0001

Cleveland, Timothy 1997. *Trying without Willing: An Essay in the Philosophy of Mind*. Aldershot, Ashgate. https://doi.org/10.4324/9781315235561

Collins, Arthur W. 1984. Action, Causality, and Teleological Explanation. *Midwest Studies in Philosophy*. 9. 345–369. https://doi.org/10.1111/j.1475-4975.1984.tb00067.x

Davidson, Donald 1971/1980. Agency. In Davidson 1980. 43–61. https://doi.org/10.1093/0199246270.003.0003

Davidson, Donald 1980. *Essays on Actions and Events*. Oxford, Oxford University Press. https://doi.org/10.1093/0199246270.001.0001

Davidson, Donald 1987/2004. Problems in the Explanation of Action. In his *Problems of Rationality*. Oxford, Clarendon Press. 101–116. https://doi.org/10.1093/0198237545.003.0007

Davis, Wayne 2010. The Causal Theories of Action. In Timothy O'Connor – Constantine Sandis (Eds.) 2010. 32–39.

Enç, Berent 2003. *How We Act: Causes, Reasons and Intentions*. Oxford, Oxford University Press. https://doi.org/10.1093/0199256020.001.0001

Gergely, György – Gergely Csibra 1998. The Teleological Origins of Mentalistic Action Explanations: A Developmental Hypothesis. *Developmental Science*. 1. 255–259. https://doi.org/10.1111/1467-7687.00039

Gergely, György – Gergely Csibra 2003. Teleological Reasoning in Infancy: The Naïve Theory of Rational Action. *TRENDS in Cognitive Sciences*. 7. 287–292. https://doi.org/10.1016/s0010-0277(97)00004-8

Gergely, György – Gergely Csibra 2007. Obsessed with Goals: Functions and Mechanisms of Teleological Interpretation of Actions in Humans. *Acta Psychologica*. 124. 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

Ginet, Carl 1990. *On Action*. Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9781139173780

Goldman, Alvin 1970. *A Theory of Human Action*. Englewood Cliffs/NJ, Prentice-Hall. https://doi.org/10.1515/9781400868971

Hawthorne, John – Nolan, Daniel 2006. What Would Teleological Causation Be? In John Hawthorne. *Metaphysical Essays*. Oxford, Oxford University Press. 265–284. https://doi.org/10.1093/acprof:oso/9780199291236.003.0015

Hornsby, Jennifer 1980. *Actions*. London, Routledge and Kegan Paul.

Hornsby, Jennifer 2004. Agency and Actions. In John Hyman – Helen Steward (Eds.) *Agency and Action*. Cambridge, Cambridge University Press. 1–24. https://doi.org/10.1017/cbo9780511550843.002

Hornsby, Jennifer 2010. Trying to Act. In Timothy O'Connor – Constantine Sandis (Eds.) 2010. 18–25. https://doi.org/10.1002/9781444323528.ch3

Huoranszki, Ferenc 2022. *The Metaphysics of Contingency: A Theory of Objects' Abilities and Dispositions*. London, Bloomsbury Academic. https://doi.org/10.5040/9781350277175

Hyman, John 2015. *Action, Knowledge, and Will*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198735779.001.0001

James, William 1983. *The Principles of Psychology*. Cambridge/MA, Harvard University Press.

Johnson, Monte R. 2005. *Aristotle on Teleology*. Oxford, Clarendon Press. https://doi.org/10.1093/0199285306.001.0001

Lennox, James G. 1985. Plato's Unnatural Teleology. In Dominic J. O'Meara (Ed.) *Platonic Investigations*. Washington D.C., Catholic University of America Press. 195–218. https://doi.org/10.2307/j.ctv176kb.12

Lowe, E. Jonathan 2000. *An Introduction to the Philosophy of Mind*. Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9780511801471

Lowe, E. Jonathan 2009. *Personal Agency*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217144.001.0001

Malcolm, Norman 1968. The Conceivability of Mechanism. *Philosophical Review*. 77. 45–72. https://doi.org/10.2307/2183182

McCann, Hugh 1974. Volition and Basic Actions. *Philosophical Review*. 83. 451–473. https://doi.org/10.2307/2183915

McGuinn, Colin 1982. *The Character of Mind*. Oxford, Oxford University Press.

McLaughlin, Brian 2012. Why Rationalization Is Not a Species of Causal Explanation. In Arto Laitinen – Constantine Sandis – Giuseppina D'Oro (Eds.) *Reasons and Causes: Causalism and anti-Causalism in the Philosophy of Action*. London, Palgrave McMillan. 97–123.

Mele, Alfred 1992. *Springs of Action*. New York, Oxford University Press. https://doi.org/10.1093/oso/9780195071146.001.0001

Mele, Alfred (Ed.) 1997. *Philosophy of Action*. Oxford, Oxford University Press.

Moravcsik, Julius 1974. Aristotle on Adequate Explanation. *Synthese*. 28. 3–17. https://doi.org/10.1007/bf00869493

O'Connor, Timothy 2000. *Persons and Causes*. New York, Oxford University Press.

O'Connor, Timothy – Constantine Sandis (Eds.) 2010. *A Companion to the Philosophy of Action*. Malden/MA, Blackwell. https://doi.org/10.1002/9781444323528

O'Shaughnessy, Brian 1973. Trying (as the mental 'pineal gland'). *Journal of Philosophy*. 70. 365–386. (Reprinted in Mele 1997.)

Plato 1997. *Phaedo*. In John M. Cooper (Ed.) *Plato: Complete Works*. Cambridge, Hackett. 49–100. https://doi.org/10.1017/s0009840x98410035

Ryle, Gilbert 1949. *The Concept of Mind*. London, Hutchison & Co.

Searle, John 1983. *Intentionality*. Cambridge, Cambridge University Press.

Searle, John 2001. *Rationality in Action*. Cambridge/MA, MIT Press.

Sehon, Scott 2005. *Teleological Realism: Mind, Agency, and Explanation*. Cambridge/MA, The MIT Press.

Smith, Michael 2010. The Standard Story of Action: An Exchange (1). In Jesus H. Aguilar – Andrei A. Buckareff (Eds.) *Causing Human Actions*. Cambridge/MA, The MIT Press. 45–55. https://doi.org/10.7551/mitpress/8614.003.0004

Sorabji, Richard 1980. *Necessity, Cause, and Blame*. Ithaca/NY, Cornell University Press.

Steward, Helen 2012. *A Metaphysics for Freedom*. Oxford, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199552054.001.0001

Strawson, Peter 1985. Causation and Explanation. In Bruce Vermazen – Merrill B. Hintikka (ed.) *Essays on Davidson: Actions and Events*. Cambridge/MA, The MIT Press. 115–135. https://doi.org/10.1093/acprof:oso/9780198751182.003.0009

Taylor, Richard 1966. *Action and Purpose*. Upper Sadle River/NJ, Prentice-Hall.

Thompson, Michael 2008. Naïve Action Theory. In his *Life and Action*. Cambridge/MA, Harvard University Press. 85–146. https://doi.org/10.4159/9780674033962

von Wright, Georg H. 1971. *Explanation and Understanding*. Ithaca/NY, Cornell University Press.

Watson, Gary 1982. Review of Hornsby's *Actions*. *Journal of Philosophy*. 79. 464–469.

Wiland, Eric 2012. *Reasons*. New York, Continuum.

Wilson, George M. 1989. *The Intentionality of Human Action*. Stanford/CA, California University Press.

Wittgenstein, Ludwig 1953. *Philosophical Investigations*. Trans. G. M. Elisabeth Anscombe. Oxford, Blackwell.

Ayumu Tamura

The Role of Experience in Descartes' Metaphysics

Analyzing the Difference Between *Intuitus*, *Intelligentia,* and *Experientia*[*]

In *Rules for the Direction of the Mind*,[1] Descartes defined intuition *(intuitio)* as "the conception of a clear and attentive mind, which is so easy and distinct that there can be no room for doubt about what we are understanding" (*Reg.*, AT-X, 368), and deduction *(deductio)* as "the inference of something as following necessarily from some other propositions which are known with certainty" (*ibid.*, 369). However, in addition to intuition, experience *(experientia)* is also presented as the opposite of deduction. Descartes states the following in Rule II:

> […] we should bear in mind that there are two ways of arriving at a knowledge of things – through experience and through deduction. (*ibid., AT-X, 365.)*

In this, experience and deduction are juxtaposed to arrive at a knowledge of things, but there is a difference in the credibility of cognition obtained by the two: "[W]hile our experiences of things are often deceptive, the deduction or pure inference of one thing from another can never be performed wrongly by an intellect which is in the least degree rational" (*ibid.*). However, this does not mean that one should completely abandon experience as a means of cognizing things. Although experience can often be wrong, it does not mean that it never gives any definite knowledge. It still depends on the type of experience. Descartes argues in Rule VIII that "it is possible to have experiential knowledge which is certain only of things which are entirely simple and absolute" (*ibid.*, 394).

[*] This study was supported by JSPS KAKENHI Grant Number 20K21950. I would like to thank Editage (www.editage.com) for English language editing. This work contains the fruit of my articles published in Japanese: Tamura (2018; 2019).

[1] For quotations from and references to Descartes, see René Descartes, *Œuvres de Descartes*, eds. Charles Adam and Paul Tannery, 11 vols. (Paris, Vrin 1964–1974), abbreviated as AT and shown in the order of conventional abbreviations, volume numbers (Roman numerals), and page numbers (Arabic numerals). I refer to the following translations: Descartes 1985a; Descartes 1985b. I make changes to them as necessary. All emphases in the quotations are by the author.

As is well known, the *Rules* citing experience, intuition, and deduction to arrive at a knowledge of things is an unfinished work written before Descartes develops the systematic idea of metaphysics. However, it is not difficult to assume that experience has an essential function in his philosophical scheme from the fact that experience is continuously used in the *Meditations*, the *Principles*, the *Conversation with Burman*, and the *Search for Truth*,[2] even though deduction and intuition are no longer thematically treated after the *Rules*.[3] Some researchers have thematically discussed what Descartes meant by "experience". For example, Clarke (1976) is the first to comprehensively treat the Cartesian experience in his study where he divided the concept broadly into two. The first is an experience as "a kind of common sense wisdom". This pertains to the ability to deal with a wide variety of environments and customs and not something purely intellectual or sensory. The second type is an experience that concerns various processes of cognition such as thought, intuition, sensation, observation, and verification. The reason for the distinction is that although it is possible for some people not to have the experience of the former (i.e., common sense wisdom), it is impossible for any human not to have the experience of the latter (i.e., thought, intuition, sensation, observation, etc.).

Clarke's study has some significance as basic research on the issue. Although the Cartesian experience can be classified this way, however, scholars must create more rigorous discussions on how the latter kind of experience relates to similar concepts such as intuition and understanding (*intelligentia*). Since this study, most other research in English-speaking countries has focused on Descartes' experience in natural science or "experiment".[4] In French-speaking countries, in contrast, Grimaldi, many years ago, and Guenancia and Kambouchner, in recent years, mentioned experience in Descartes' metaphysics. However, there are still many points that scholars must investigate, as mentioned in the next section.

In this paper, I intend to explore what Descartes meant by the term "experience" in the context of metaphysics. To be concrete, I first compare Descartes with earlier philosophers and clarify that Descartes' use of the term "experience" has characteristics that were not recognized earlier (Section 1). I then

[2] Experience is used in the themes of the *cogito*, God, and free will that underlie his metaphysics (*Med.*, AT-VII, 49; *ibid.*, 56; *2ae Resp.*, AT-VII, 140; *5ae Resp.*, AT-VII, 358; *6ae Resp.*, AT-VII, 427; *P.Ph.*, AT-VIII, 19-20; *ibid.*, 33; *Ent. Burm.*, AT-V, 147; *ibid.*, AT-V, 163; *R.V.*, AT-X. 524).

[3] See Garber 1992. 56–57.

[4] There is a section for "experiment" but none for "experience" in *The Cambridge Descartes Lexicon* (Nolan 2016), which introduces the latest findings of research. This term is described together with experiment as "experience (experiment)" in the *Historical Dictionary of Descartes and Cartesian Philosophy* (Ariew 2015). Alanen (2003. 266–267 [n. 21]) briefly mentions Descartes' notion of experience, but she does not go beyond the framework of Clarke's research.

clarify what the role of experience in Descartes is, while examining the validity of previous studies that equate Descartes' experience with intuition or understanding (sections 2 and 3).

I. THE PECULIARITY OF DESCARTES' USE OF EXPERIENCE

1. Before Descartes

The concept of experience has been an important part of philosophy since the ancient times. We can look at Aristotle as an example. Setting aside the validity of Heinemann's view (1941. 562) that Aristotle is the first philosopher who defined experience,[5] it is at least clear that he was one of the earliest philosophers who emphasized on the method of experience in academic knowledge. Aristotle argued on experience as follows: "[F]rom memory experience is produced in men; for the several memories of the same thing produce finally the capacity for a single experience. And […] science and art come to men through experience […]" (Aristotle 2007. 2205). In other words, the perceptions given by the senses accumulate as memories and are appropriately categorized and sublimated into one experience of the same thing. Knowledge and skills arise from the experience thereafter. In the words of Gregorić and Grgić, the Aristotelian experience can be described as something that "fills a wide gap between the non-rational cognitive capacities of perception and memory on the one side, and the rational cognitive dispositions of art and science on the other side" (Gregorić 2006. 2).

Medieval philosophy was strongly influenced by him – "experience" was Aristotelian. According to Albert the Great (Albert 1960. 13), experience is the cognition about the individual things received from repeated memories ("*experientia est cognitio singularium ex multiplicatis accepta memoriis*"), and in order to have knowledge through experience, there must be three separate mental events: (1) an impression of something, (2) an impression of another thing similar to it, and (3) an act of taking the two preceding impressions, at least one of which is recalled from memory (King 2003. 8). Thomas Aquinas also states that "we ourselves have experience when we know singular things through sensation" (Aquinas 2018. Prima Pars, q. 54, art. 5) and that "[one] has memory and experience of [the particulars] through the sensory power" (Aquinas 2018. Prima Pars, I, q. 117, art. 1).[6] For Aquinas, experience/to experience is *something that arises from multiple memories/to cognize individual things through the senses*, which is based

[5] According to Gregorić and Grgić (2006. 1–30), Aristotle did not define "experience."

[6] The original word here is *experimentum*, but it was used synonymously with *experientia* at least until the late Middle Ages (Park 2011. 38 [n. 4]).

on Aristotle's view. As "experience" originated in the senses, it was never the chief method in metaphysics in the medieval era. This point is evident in Duns Scotus' *Ordinatio*. He writes thus:

> It must be noted, further, that sometimes experience concerns [not a principle itself, as was the case in the preceding paragraph, but rather a] conclusion, as, for example, that the moon is at times eclipsed. Then one assumes that the conclusion holds and investigates the cause of such a conclusion by means of an analysis. And sometimes an empirical conclusion *(conclusione experta)* leads to principles that are known from their terms. In that case, one can on the basis of such principles known through their terms get more certain knowledge of the conclusion that was initially only known empirically *(secundum experientiam)*. This is an instance of the first category of certain knowledge, for it is deduced from a principle known per se. For example, it is known per se that "when something opaque is put between a light source and a clearly visible body, it prevents the propagation of light to the body." If, then, it is found out by analysis that the earth is such a body put between the sun and the moon, knowledge [of the eclipse] will be had with maximal certainty based on a demonstration giving the reason or the cause. The conclusion will not just rest on experience, as was the case before the [explanatory] principle was found. (Scotus 2016. 125)

According to Scotus, if a proposition placed in the position of the conclusion of a syllogism is known in advance by experience, it can be considered a sound argument as a whole by exploring its principle retroactively from the conclusion. That is, on the one hand, the presupposed self-evident principle is obtained by returning from the empirical proposition as a conclusion. On the other hand, the self-evident principle obtained *a posteriori* guarantees the certainty of the empirical proposition. It follows from this that there was a difference in the certainty between what is known by an experience and by [the deduction from] the principle even if the two pertain to the same thing.[7] That is, empirical knowledge is considered inferior to deductive knowledge that is derived from principle.

The understanding that experience arises from the senses and memories is also recognized in Michel de Montaigne's *Essays*, which Descartes loved to read. In Chapter 13, titled *Of Experience* in the third volume, Montaigne writes thus:

[7] According to Descartes, such a method is useless for discovering the truth. "But on further examination I observed with regard to logic that syllogisms and most of its other techniques are of less use for learning things than for explaining to others the things one already knows or even, as in the art of Lully, for speaking without judgement about matters of which one is ignorant." (*D.M.*, AT-VI. 17; *cf. Principes*, AT-IX. 13)

> There is no desire more natural than the desire for knowledge. We try all the ways that can lead us to it. When reason fails us, we use experience –
>
> Experience, by example led,
> By varied trials art has bred
> MANILIUS
>
> – which is a weaker and less dignified means. But truth is so great a thing that we must not disdain any medium that will lead us to it. (Montaigne 1965. 815)

Here, Montagne obviously inherited from Aristotle the idea that experience arises from memories, and inherited from medieval philosophers such as Scotus, the idea that experience is inferior to reason. For him, experience is most useful in medicine. He writes that "Experience is really on its own dunghill in the subject of medicine, where reason yields it the whole field" (*ibid.* 826). He also claims that "[M]edicine professes always to have experience as the touchstone for its workings" (*ibid.* 827). However, Montaigne did not rely on the method of experience while discussing metaphysical matters.

We can fully surmise that experience did not play a major role in metaphysics by referring to various Latin dictionaries compiled in the 16th and 17th centuries. According to Micraelius' *Lexicon philosophicum*, published shortly after Descartes died,

> "Experience is the general knowledge constructed of a number of individual and [mutually] similar things (*Experientia est ex pluribus singularibus cognatis scientia universalis exstructa*)" (Micraelius 1653. 417). In *Lexicon philologicum*, Martini writes that "[Experience] is, first, a sense, second, an observation, third, an experience and fourth, an induction. Therefore, it is also the general rule [derived by an induction] (*Primo est sensus, secundo observatio, tertio experientia, quarto inductio, hinc generalis regula*)" (Martini 1655. art. *Experientia*). According to Chauvin's *Lexicon rationales*, which was compiled in the second half of the 17th century, "Experience is a kind of cognition which is not taught by anyone but arises from a practice or a habit. Only in natural science each person has experience, and he or she necessarily has experience. This is because reason without experience is equal to a swaying ship without a person steering (*Experientia est quaedam cognitio nullo docente, per usum contingens. In Physicis tantum obtinet, sed & necessario obtinet; est enim ratio sine experientia velut navis sine rectore fluctuans.*)" (Chauvin 1692. art. *Experientia*).

These dictionaries have a description from the perspective of natural science but not metaphysics, and Goclenius' *Lexicon philosophicum* (1590), which was compiled in the latter half of the 16th century, does not have a section on experience in the first place. We can say that these dictionaries represent aspects of experience as a philosophical concept that continued for a hundred years that centers on Descartes' death (1650).

2. Descartes: Introducing Experience into Metaphysics

Descartes used the term "experience" heavily in the field of metaphysics. The usage is very different from that of the pre-Descartes tradition identified in the previous section. Its most prominent characteristic is that the objects of experience include external things perceived through the senses and internal things that appear in the mind. These include thought, free will, the union of mind and body, and so on. I summarize some of the main examples in a table.

<table>
<tr><td colspan="3">Experientia/experimentum/expérience (including the verb experiri/expérimenter)</td></tr>
<tr><td colspan="2">Experience concerning the mind</td><td>Experience concerning the union of mind and body</td></tr>
<tr><td>Metaphysical</td><td>Others</td><td>/</td></tr>
<tr><td>Med., AT-VII:
38: "I know by experience that these ideas do not depend on my will, and hence that they do not depend simply on me."
53: "I know by experience that there is in me a faculty of judgement […]."
55: "[T]here is no call to doubt his existence if I happen to experience that there are other instances where I do not grasp why or how certain things were made by him."
57: "It is only the will, or freedom of choice, which I experience within me to be so great that the idea of any greater faculty is beyond my grasp […]."

Resp., AT-VII:
140: "[Someone who says, 'I am thinking, therefore I am'] learns it from experiencing in his own case that it is impossible that he should think without existing."
191: "On the question of our freedom, I made no assumptions beyond what we all experience within ourselves."
358: "[T]he mind, when engaged in private meditation, can experience its own thinking but cannot have any experience to establish whether the brutes think or not […]."</td><td>Med., AT-VII:
47: "I am now experiencing a gradual increase in my knowledge […]."
52: "[E]xperience tells us that this same contemplation, albeit much less perfect, enables us to know the greatest joy of which we are capable in this life."
54: "But when I turn back to myself, I know by experience that I am prone to countless errors."
59: "My experience in the last few days confirms this: the mere fact that I found that all my previous beliefs were in some sense open to doubt was enough to turn my absolutely confident belief in their truth into the supposition that they were wholly false."
62: "Admittedly, I [experience] a certain weakness in me, in that I am unable to keep my attention fixed on one and the same item of knowledge at all times […]."
71: "The conclusion that material things exist is also suggested by the faculty of imagination, which I [experience] that I use when I turn my mind to material things."</td><td>Resp., AT-VII:
228: "[T]he fact that the mind is closely conjoined with the body, which we experience constantly through our senses […]."

P.Ph., AT-VIII:
23: "But we also experience within ourselves certain other things which must not be referred either to the mind alone or to the body alone."

Ent. Burm., AT-V:
163: "[How the soul can be affected by the body and vice versa, when their natures are completely different] is very difficult to explain; but here our experience is sufficient, since it is so clear on this point that it just cannot be gainsaid. This is evident in the case of the passions, and so on."</td></tr>
</table>

Experientia/*experimentum*/expérience (including the verb *experiri*/expérimenter)		
Experience concerning the mind		*Experience concerning the union of mind and body*
Metaphysical	*Others*	/
427: "We cannot fail constantly to experience within ourselves that we are thinking." ***P.Ph.*, AT-VIII:** **6:** "But whoever turns out to have created us, and however powerful and however deceitful he may be, in the meantime we nonetheless experience within us the kind of freedom which enables us always to refrain from believing things which are not completely certain and thoroughly examined." **17:** "All the modes of thinking that we experience within ourselves can be brought under two general headings […]." ***Ent. Burm.*, AT-V:** **147:** "[…] I am attending only to what I experience within myself – for example 'I am thinking, therefore I exist'. I do not pay attention in the same way to the general notion 'whatever thinks exists'." **159:** "Let everyone just go down deep into himself and find out [by experience] whether or not he has a perfect and absolute will, and whether he can conceive of anything which surpasses him in freedom of the will. I am sure everyone will find [by experience] that it is as I say." ***R.V.*, AT-X:** **524:** "[W]hat convinces us of [thought, existence and certainty] is simply our own experience or awareness – that awareness or internal testimony which everyone experiences within himself when he ponders on such matters."	**75:** "For my experience was that these ideas came to me quite without my consent, so that I could not have sensory awareness of any object, even if I wanted to, unless it was present to my sense organs; and I could not avoid having sensory awareness of it when it was present." ***Resp.*, AT-VII:** **230:** "[O]ur own experience reliably informs us that [the sight of the impending fall reaches the brain and sends the animal spirits into the nerves in the manner necessary to produce this movement even without any mental volition, just as it would be produced in a machine] […]." **358:** "But when the imagination is less intense, we often have the experience of understanding something quite apart from the imagination." **365:** "[E]ach of us knows by his own experience quite well that he has this sort of understanding of the infinite […]." **429:** "I [do not experience] so much divine grace within me that I feel a vocation for such sacred studies." ***P.Ph.*, AT-VIII:** **16:** "[…] the cause of the errors to which we know by experience that we are prone." ***Ent. Burm.*, AT-V:** **148:** "I have nothing to say on the subject of memory. Everyone should [know by experience] whether he is good at remembering."	

From these examples, we see that experience is crucial to Descartes in establishing the certainty of metaphysical knowledge. We can infer by referring to the French versions of the *Meditations* and the *Principles* compiled by third parties that his use of such "experience" was peculiar for his time. The word "experience" in the original Latin text is replaced by a completely different word in the French versions of the *Meditations* and the *Principles* as it was deemed appropriate: (1) "*Itaque debeo nunc interrogare me ipsum, an habeam aliquam vim per quam possim efficere ut ego ille, qui jam sum, paulo post etiam sim futurus […]. Sed & nullam esse* **experior** *[…]*" (*Med.*, AT-VII. 49) was changed to "Il faut donc seulement ici que je m'interoge moi-même, pour savoir si je possède quelque pouvoir et quelque vertu, qui soit capable de faire en sorte que moi, qui suis maintenant, sois encore à l'avenir: […]; mais je n'en ***ressens*** aucune dans moi […]" (AT-IX-1. 39); (2) "*[…] apud se* **experiatur**, *fieri non posse ut cogitet, nisi existat*" (*2ae Resp.*, AT-VII, 140) was changed to "[…] il ***sent*** en lui-même qu'il ne se peut pas faire qu'il pense, s'il n'existe" (AT-IX-1. 110-111); (3) "*[…] omnes modi cogitandi, quos in nobis* **experimur**, *ad duos generales referri possunt […]*" (*P.Ph.*, AT-VIII, 17) was changed to "[…] toutes les façons de penser que nous ***remarquons*** en nous, peuvent être rapportées à deux générales […]" (AT-IX-2. 39); (4) "*[…] hanc in nobis libertatem esse* **experiebamur** *[…]*" (*P.Ph.*, AT-VIII, 20) was changed to "[…] nous ***apercevions*** en nous une liberté […]" (AT-IX-2. 41); (5) "*[…] nec ullam similitudinem intelligere possimus, inter colorem quem supponimus esse in objectis, & illum quem* **experimur** *esse in sensu […]*" (*P.Ph.*, AT-VIII, 34) was changed to :[…] notre raison ne nous fasse apercevoir aucune ressemblance entre la couleur que nous supposons être en cet objet & celle qui ***est*** en notre sens […]" (AT-IX-2. 58). We can say that such changes suggest that during that period, the Cartesian use of "*experientia*" seemed alien to the French translators.

3. From Montaigne to Descartes

Unlike earlier philosophers, what led Descartes to use the word "experience" heavily in metaphysics? One possible interpretation is that Descartes attempted to apply the natural scientific method of experience (experiment) to metaphysics. Montaigne placed utmost importance on experience in the field of medicine. He writes thus:

> It is reasonable that [a doctor] should catch the pox if he wants to know how to treat it. Truly I should trust such a man. For the others guide us like the man who paints seas, reefs and ports while sitting at his table, and sails the model of a ship there in complete safety. (Montaigne 1965. 827)

It is necessary to know what illness X is first to be able to know the cure for it. However, Montaigne explains that to know what X is, it is not enough to observe a person suffering from X *from the outside*. The doctor must observe the situation caused by X *within himself*. Otherwise, there will only be theoretical arguments about X. Descartes' metaphysics also reflected such an argument. He writes thus:

> [T]he only way we can learn such things (= doubt and thought) is by ourselves: what convinces us of them is simply our own experience or awareness – that awareness or internal testimony which everyone experiences within himself when he ponders on such matters. Thus it would be pointless trying to define, for someone totally blind, what it is to be white: in order to know what that is, all that is needed is to have one's eyes open and to see white. In the same way, in order to know what doubt and thought are, all one need do is to doubt or to think. That tells us all it is possible to know about them, and explains more about them than even the most precise definitions. (*R.V., AT-X, 524)*

Here, Descartes seems to have kept in mind Montaigne's claim. As one needs to actually get syphilis to learn how to cure it, and to open his or her eyes and actually see what is white to know the color, one must actually think and doubt to know what thought and doubt are. Otherwise, one may end up just playing with theoretical arguments on thought and doubt. Descartes writes, "I have often noticed that philosophers make the mistake of employing logical definitions in an attempt to explain what was already very simple and self-evident; the result is that they only make matters more obscure" (*P.Ph.*, AT-VIII, 8). In other words, Descartes sought to emphasize experience not only in the fields of natural science and medicine, but also in metaphysics. Therefore, Descartes' metaphysics is different from that of earlier philosophers and seems to be regarded *experiential metaphysics*.[8]

Based on the above-mentioned considerations, we can see that to understand Descartes' metaphysics, it is essential to clarify his concept of experience. How then has previous research treated this concept? Scholars have tended to identify experience with other intellectual acts such as "intuition (*intuitus*)" and "understanding (*intelligentia*)." Hamelin (1921. 75) argues that there is a well-defined experience that covers distinct notions in the Cartesian system, and such an experience is "a kind of intuition." Grimaldi (1978. 101) also claims that intuition is an experience of simple nature by intelligence, therefore, absolute certainty

[8] The connection between Montaigne's *Of Experience* (in the *Essays*) with Descartes' *Search for Truth* is unverifiable because there is no direct evidence showing that Descartes kept Montaigne in mind while writing the *Search for Truth*. However, my interpretation does not seem invalid considering that Descartes was hugely influenced by the writings of Montaigne.

is possible only in that intuition and one may define it as the "metaphysical experience of the truth" (*l'expérience métaphysique de la vérité*). Guenancia (2009. 64) argues that the experience of a clearly present object is nothing more than a certain perception of a thing, and its certainty is an epistemological expression of a direct experience that the mind obtains with things through intuition. Kambouchner (2015. 128) argues that every piece of evidence and intellectual intuition as provided in the *Rules* is the full experience of an object. This is an experience of the necessity that things are a certain way, or an experience of the impossibility that things cannot be any other way. Therefore, *intelligere* and *experiri* are synonymous.

On my reading, however, equating experience with intuition or understanding seems to be impossible considering Descartes' meditative transition from his early *Rules* to the *Meditations* and beyond. It would only overlook the unique role included in experience. In the next section, I examine the unique function by comparing and contrasting experience with both intuition and understanding while critically examining previous research.

II. *INTUITUS* AND *EXPERIENTIA*

In this section, I examine the relationship between experience and intuition. At first glance, it seems possible to interpret that, for Descartes, experience is synonymous with intuition or the latter is a subdivision of the former, because these concepts are presented in a paired relationship of *deduction-experience* and *deduction-intuition*. This being so, it follows that experience and intuition have the same function and degree of certainty in metaphysics. On my reading, however, such an interpretation cannot be established unless we restrict the discussion to the *Rules*.[9] This is because what was claimed to be definitely known by intuition in the *Rules* is doubted in the *Meditations*: "that [one] exists, that [one] is thinking, that a triangle is bounded by just three lines, and a sphere by a single surface, and the like" (*Reg.*, AT-X, 368) are considered known by intuition in the *Rules*, but among them, the certainty of mathematical and geometrical knowledge is rejected by the methodological doubt in the *Meditations*: he states, "[…] since I sometimes believe that others go astray in cases where they think they have the most perfect knowledge, may I not similarly go wrong every time I add two and three or count the sides of a square, or in some even simpler matter, if

[9] The eight notes Grimaldi adds to the section where this issue is discussed in his book all refer to the *Rules* (Grimaldi 1978. 100–101). Hamelin's argument that "*experientia*" and "*intuitus*" are used as synonyms concerns the *Rules* and not the *Meditations* or the *Search for Truth* (Hamelin 1921. 75). Guenancia, on the contrary, seems to identify the metaphysical experience that can guarantee the truth of the *cogito* as the intuition in the *Rules* (Guenancia 2009. 65).

that is imaginable?" (*Med.*, AT-VII, 21). In the Second Reply, the *Search for Truth* and the *Conversation with Burman*, Descartes claims that one needs "to experience" one's own thoughts and existence rather than "have an intuition" of them.

If we, ignoring these points, equate the intuitions that include the mathematical knowledge mentioned in the *Rules* with the experience mentioned in Descartes' later works that discuss metaphysics, it follows that the level of the certainty of things that are said to be known by experience in metaphysics (especially the certainty of the *cogito*) is equal to that of the mathematical knowledge that will be rejected by the methodological doubt. Therefore, it was impossible for Descartes to maintain the certainty of experience *at a metaphysical level* when he said that "the mind […] can experience its own thinking" (*5ae Resp.*, AT-VII, 358) and that "[one] learns it from experiencing in [one's] own case that it is impossible that [one] should think without existing" (*2ae Resp.*, AT-VII, 140–141).[10] After the *Rules*, he rarely used intuition as a specialized philosophical concept[11]: there are seven examples of "intuition" in the *Meditations*, two of which are used in the quite ordinary sense of "staring;" the other five examples are used as those that can be compared to sight and are closely related to imaginations: "When I imagine a triangle, for example, I do not merely understand that it is a figure bounded by three lines, but I also see the three lines with my mind's eye as if they were present before me. This is what I call imagining" (*Med.*, AT-VII, 72).

Let us also consider the following points in the Second Reply:

> And when we become aware that we are thinking things, this is a primary notion that is not derived by means of any syllogism. When someone says 'I am thinking, therefore I am, or I exist', he does not deduce existence from thought by means of a syllogism, but recognizes it as something self-evident by a simple intuition of the mind. This is clear from the fact that if he were deducing it by means of a syllogism, he would have to have had previous knowledge of the major premiss 'Everything which thinks is, or exists'; yet in fact he learns it from experiencing in his own case that it is impossible that he should think without existing. It is in the nature of our mind to construct general propositions on the basis of our knowledge of particular ones. (*2ae Resp., AT-VII, 140–141)*

The first impression is that Descartes adopted intuition as one of the valid methods in metaphysics. He does write that "I am thinking, therefore I am" is known "by a simple intuition of the mind." On my reading, however, the "intuition" mentioned here does not have any academically special meaning.

[10] The object of intuition in the *Rules* is not "I think, therefore I am," but "I think" and "I exist," separately. Descartes made no mention of the connection between thought and existence in the *Rules* and did not argue for the existence of "I" as an entity distinguished from the body.

[11] Marion (1977. 295–302) argued that "*intuitus*" should be translated as "regard."

The phrase "recognizes it as something self-evident by a simple intuition of the mind" does not mean more than recognizing without the mediation of a major premise. Rather, the emphasis in this section is on the explanation that "he learns it (=existence) from experiencing in his own case that it is impossible that he should think without existing." In other words, Descartes acknowledged that in order to cognize one's own existence "by a simple intuition of the mind," the major premise "Everything which thinks is" is not necessary, but "experiencing in [one's] own case that it is impossible that [one] should think without existing" is necessary. The intuition here is different from the intuition in the *Rules*, in that the latter is just "tak[ing] in at one glance" (*Reg.*, AT-X, 370) propositions and the link with plural propositions and the former is recognizing particular things[12] through the practice where one achieves cognition within oneself[13] (*2ae Resp.*, AT-VII, 141). Whereas Descartes presented arguments in the *Meditations* and the Replies, he realized that the intuition he mentioned in the *Rules* could no longer serve as a method for his metaphysical investigation.[14] Thus, he began to use the method of experience instead of intuition when he discussed metaphysical subjects, such as the *cogito* and free will.

III. *INTELLIGENTIA* AND *EXPERIENTIA*

Next, I examine the relationship between experience and understanding. As mentioned earlier, Kambouchner asserted that experience and understanding are the same thing. However, is such an interpretation appropriate? Stating from the conclusion, there is a difference between both concepts in terms of the object's actual *presence*, and this presence is the most distinctive attribute of experience.

The presence of an object of perception has a compelling force on the subject in this sentence: "I could not have sensory awareness of any object, even if I wanted to, unless it was present to my sense organs; and I could not avoid having sensory awareness of it when it was present" (*Med.*, AT-VII, 75). Understanding

[12] According to Rosenthal (1986. 422), "understanding some propositional content does not typically pin down the reference of whatever token-reflexive components are involved. Understanding 'Theaetetus is sitting' does not determine the time that the present tense refers to; understanding 'He gives it to her' does not suffice to pick out any particular people or gift. To pin down reference in such cases, one typically needs more than an act of understanding." We may assume that because Descartes was aware of this point, he abandoned the intuition defined in the *Rules*. It was no more than simply grasping the content of a proposition.

[13] The experience that *fieri non posse ut cogitet, nisi existat/il ne se peut pas faire qu'il pense, s'il n'existe* (AT-VII, 140/AT-IX. 111) is differentiated from the impersonal general proposition "In order to think it is necessary to exist (*pour penser, il faut être*)". The former has a personal noun (third person singular, i.e., the one who says, "I think, therefore I am").

[14] See Curley 1978. 38; Garber 1992. 56–57.

does not imply the presence of the object from this sentence: "When I imagine a triangle, for example, I do not merely understand that it is a figure bounded by three lines, but at the same time I also see the three lines with my mind's eye as if they were present before me" (*Med.*, AT-VII, 72). In other words, understanding alone cannot lead one to grasp the object as a *thing that is present*. For example, a geometrically perfect triangle can be understood through analytical geometry (i.e., by converting it into a mathematical formula), but it will no longer be the triangle itself (as a shape). The object of understanding is no longer present as it was.[15] In contrast, the object of experience is things that are present. Experience is the act of *making an object present*. To clarify this, let us consider the concept of thought, which Descartes most emphasized as what can be known by experience.[16] According to the *Search for Truth*, we do not have to "rack our brains trying to find the 'proximate genus' and the 'essential differentia' which go to make up their true definition" (*R.V.*, AT-X, 523); rather, we can know it by "our own experience" (*ibid.*, 524). Descartes compared this as follows: "[I]t would be pointless trying to define, for someone totally blind, what it is to be white: in order to know what that is, all that is needed is to have one's eyes open and to see white" (*ibid.*). Just as one needs to actually see something white to know what white is, one also needs to actually think to know what thought is. "The mind […] can experience its own thinking" (5*ae Resp.*, AT-VII, 358) means that the mind "is actually thinking" (*2ae Resp.*, AT-VII, 151), and for the mind to actually think is nothing but for the mind to make a thought about something present in itself. Descartes considered thought "what [one] *cannot fail to experience* within [oneself]" (*6ae Resp.*, AT-VII, 427) precisely because it is present to the mind. The mind cannot resist the compelling force of the presence of thought and cannot help but accept that the thought is in the mind, as long as it is present. Descartes also argued the following: "I know by experience that there is in me a faculty of judgement" (*Med.*, AT-VII, 53) and "I know by experience that [the

[15] Let us compare the following texts:
1. "[W]e understand [the mind] to exist without the body" (*4ae Resp.*, AT-VII, 227).
2. "We know by experience that our minds are so closely joined to our bodies" (*À X****, AT-III, 423–424).

Descartes claims that "we can understand that the mind exists without the body" and explains that "I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God" (*Med.*, AT-VII, 78). The mind cannot experience that the mind and body are separate entities and not interdependent even though the mind can understand it by the intellect. Descartes says that the mind can understand that it exists without the body. However, he does not say that the mind can experience it. What the mind experiences is that it is given various sensations and emotions by the body and that the will of the mind causes physical movements (such as walking and bending/stretching of arms).

[16] "[T]he mind, when engaged in private meditation, can experience its own thinking" (*5ae Resp.*, AT-VII, 358; *cf. 6ae Resp.*, AT-VII, 427; *Ent. Burm.*, AT-V, 147; *R.V.*, AT-X, 524).

will] is not restricted in any way. Indeed, I think it is very noteworthy that there is nothing else in me which is so perfect and so great that the possibility of a further increase in its perfection or greatness is beyond my understanding" (*ibid.*, 56). The experience mentioned here has a role to make present the situation where "the ability to make a judgement exists within oneself." Such is done by actually carrying out the act of judgment. Free will is presented in a way that it is not restricted by any limitations by actually making a decision. Even with the methodological doubt that denies everything, the presence of thought, a faculty of judgment, and free will can never be denied.

In addition, experience in this sense has an extremely close relationship with consciousness in modern language. When we experience that we are thinking, that there is in us a faculty of judgment, and that our own will is not restricted in any way, it is impossible for us not to be self-aware of actually carrying out such acts by ourselves in that situation. Although we are often unaware of physical acts even if they are actually being carried out (e.g., turning over during sleep), it is impossible for us not to be aware of purely non-physical acts such as thought, judgment, and decision making. Descartes uses the words *experientia* and *conscientia* interchangeably, as the following texts show:

> *Itaque debeo nunc interrogare me ipsum, an habeam aliquam vim per quam possim efficere ut ego ille, qui jam sum, paulo post etiam sim futurus : nam, cum nihil aliud sim quam res cogitans, vel saltem cum de ea tantum mei parte praecise nunc agam quae est res cogitans, si quae talis vis in me esset, ejus proculdubio* **conscius** *essem. Sed & nullam esse* **experior**, *& ex hoc ipso evidentissime cognosco me ab aliquo ente a me diverso pendere.* (*Med.*, AT-VII, 49)

> *Nam sane fieri non potest quin semper apud nosmetipsos* **experiamur** *nos cogitare ; nec proinde ex eo quod ostendatur bruta animantia omnibus suis operationibus absque ulla cogitatione fungi posse, quisquam concludet seipsum ergo etiam non aliter operari quam bruta, propter hoc scilicet quod illis cogitationem tribuerit, adeo pertinaciter adhaerebit istis verbis, homines & bruta eodem modo operantur, ut, cum illi ostendetur bruta non cogitare, malit se etiam illa sua, cujus non potest non esse sibi* **conscius**, *cogitatione exuere, quam mutare opinioionem quod ipse eodem modo ac bruta operetur.* (*6ae Resp.*, AT-VII, 427)

> *[…] libertatis autem & indifferentiae, quae in nobis est, nos ita* **conscios** *esse, ut nihil sit quod evidentius & perfectius comprehendamus. Absurdum enim esset, propterea quod non comprehendimus unam rem, quam scimus ex natura sua nobis esse debere incomprehensibilem, de alia dubitare, quam intime comprehendimus, atque apud nosmet ipsos* **experimur**. (*P.Ph.*, AT-VIII, 20)

> *[…] hocque propter aptam corporis configurationem, quam mens potest ignorare, ac etiam propter mentis cum corpore unionem, cuius sane mens* **conscia** *est ; alioquin enim ad membra movenda voluntatem suam non inclinaret.* (New line) *Quod autem mens, quae incorporea*

> *est, corpus possit impellere, nulla quidem ratiocinatio vel comparatio ab aliis rebus petita, sed certissima & evidentissima* **experientia** *quotidie nobis ostendit [...].* (*À Arnauld* [29 juillet 1648], AT-V, 222)

> *Verum his adjungo, fieri non posse, ut alia quis ratione, ac per se ipsum, ea addiscat, neque ut de iis alio modo persuasus sit, quam propria* **experientia**, *eaque* **conscientia**, *vel interno testimonio, quod in se ipso unusquisque, cum res perpendit, experitur.* (*R.V.*, AT-X, 524)

It is still disputed among scholars whether it is possible to translate Descartes' *conscientia* as "consciousness". However, we do not have to address this here. We should be cautious about considering the Cartesian *conscientia* as the philosophical term "consciousness", but it is true that this Latin word traditionally means not only conscience, as in ethical valuation, but also an awareness unrelated to ethical valuation, such as witness and testimony. According to Hennig (2007. 455–484), *conscius* means "witness" when used as a noun. The word implies that a person knows about a crime or an event because he or she was involved in it. *Conscientia* was considered a witness to an event one was involved in (Seneca writes, "When one has one's own *conscientia*, what is the problem of not having a witness?", and Quintilianus writes, "*Conscientia* is equivalent to a thousand witnesses"). It is important to note that Descartes used the word *conscientia*, which had traditionally meant witness and testimony, interchangeably with *experientia*. For Descartes, experience not only meant that the mind actually carried out purely non-physical acts such as thinking and judging. It also meant that the mind witnesses and testifies to what arises within itself from those acts. The mind may not be able to witness or testify to some physical acts (e.g., turning over during sleep). However, in the case of non-physical (i.e., purely mental) acts, such as thinking and judging, it is never possible for the mind to not be able to witness or testify to them, as the mind carries them out within itself.

IV. CONCLUSION

Based on the analysis and above-mentioned considerations, experience has unique attributes that differ from those of intuition and understanding in Descartes' philosophical system. For Descartes, experience refers to actually carrying out purely non-physical acts such as thinking and judging. It also means to witness and testify to (i.e., be conscious of) what arises in the mind from those acts.

It seems that the replacement of *intuitus* in the *Rules* with *experientia*, which connotes consciousness is inevitable from the transition of Descartes' thought. According to the *Rules*, intuition is "the conception of a clear and attentive mind, which is so easy and *distinct*" (*Reg.*, AT-X, 368). However, not everything that

arises within oneself is clear and distinct: pain and pleasure are clear but not distinct.[17] Thus, if we use intuition (and understanding) alone consistently, we will not be able to capture these things; therefore, we will not be able to reach the Sixth Meditation, which discusses the relationship between the mind and body. Descartes used experience as a means to capture our thought and existence as well as our internal senses such as clear but indistinct pain and pleasure. The mind experiences its own thoughts, existence, free will, and pain and pleasure. The mind *is conscious of* them.

Abbreviations for the Works of Descartes

Ent. Burm. = *Entretien avec Burman*
Med. = *Meditationes de prima philosophia*
P.Ph. = *Principia philosophiae*
Reg. = *Regulae ad directionem ingenii*
Resp. = *Responsiones*
R.V. = *Recherche de la vérité*

REFERENCES

Alanen, Lilli 2003. *Descartes's Concept of Mind*. Cambridge/MA, Harvard University Press. https://doi.org/10.4159/9780674020108

Albert 1960. *Alberti Magni ordinis fratrum praedicatorum Metaphysica*. Vol. 16. Ed. Berhardt Geyer. Aschendorff, Münster Westfalen.

Aquinas, Thomas 2018. *Summa Theologiae*. Trans. Alfred J. Freddoso. https://www3.nd.edu/~afreddos/summa-ranslation/TOC.htm

Aristotle 2007. *Metaphysics*. In William D. Ross (trans.) *Aristotle Organon and Other Works*. Internet Archive. https://archive.org/details/AristotleOrganon/mode/2up

Ariew, Roger (ed.) 2015. *Historical Dictionary of Descartes and Cartesian Philosophy*, 2nd edition. Lanham/MD, Rowman & Littlefield. https://doi.org/10.5860/choice.192446

Chauvin, Étienne 1692. *Lexicon rationale sive Thesaurus philosophicus*. Rotterodami: apud Petrum van der Slaart.

Clarke, Desmond M. 1976. The Concept of Experience in Descartes' Theory of Knowledge. *Studia Leibnitiana*. 8/1. 18–39.

Curley, Edwin M. 1978. *Descartes against the Skeptics*. Oxford, Blackwell. https://doi.org/10.4159/harvard.9780674330245

de Montaigne, Michel 1965. *Complete Essays*. Trans. Donald M. Frame. Stanford, Stanford University Press.

Descartes, René 1964–1974. *Œuvres de Descartes*. Eds. Charles Adam and Paul Tannery. 11 vols. Paris, J. Vrin.

[17] "For example, when someone feels an intense pain, the perception he has of it is indeed very clear, but is not always distinct" (*P.Ph.*, AT-VIII, *22*).

Descartes, René 1985a. *The Philosophical Writings of Descartes*. Vol. 1 and 2. Trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9780511805042, https://doi.org/10.1017/CBO9780511818998

Descartes, René 1985b. *The Philosophical Writings of Descartes*. Vol. 3. Trans. John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny. Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9781107340824

Garber, Daniel 1992. *Descartes' Metaphysical Physics*. Chicago, University of Chicago Press.

Gregorić, Pavel – Filip Grgić 2006. Aristotle's Notion of Experience. *Archiv für Geschichte der Philosophie*. 88/1. 1–30. https://doi.org/10.1515/AGPH.2006.001.

Grimaldi, Nicolas 1978. *L'Expérience de la pensée dans la philosophie de Descartes*. Paris, J. Vrin.

Guenancia, Pierre 2009. *Descartes, chemin faisant*. Paris, Les Belles Lettres.

Hamelin, Octave 1921. *Le Système de Descartes*. Paris, Alcan.

Heinemann, Frederick H. 1941. The Analysis of Experience. *The Philosophical Review*. 50/6. 561–584. https://doi.org/10.2307/2180812

Hennig, Boris 2007. Cartesian *Conscientia*. *British Journal for the History of Philosophy*. 15. 455–484. https://doi.org/10.1080/09608780701444915

Kambouchner, Denis 2015. *Descartes n'a pas dit*. Paris, Les Belles Lettres.

King, Peter 2003. Two Conceptions of Experience. *Medieval Philosophy and Theology*. 11. 203–226. https://doi.org/10.1017/S1057060803000082

Marion, Jean-Luc (ed.) 1977. *Règles utiles et claires pour la direction de l'esprit en la recherche de la vérité*. La Haye, Nijhoff.

Martini, Matthias 1655. *Lexicon philologicum*. Goetzen.

Micraelius, Johann 1653. *Lexicon philosophicum terminorum philosophis usitatorum*. Freyschmid.

Nolan, Lawrence (ed.) 2016. *The Cambridge Descartes Lexicon*. Stanford/CA, Cambridge University Press. https://doi.org/10.1017/CBO9780511894695

Park, Katharine 2011. Observation in the Margins, 500–1500. In Lorraine Daston and Elizabeth Lunbeck (eds.) *Histories of Scientific Observation*. Chicago, University of Chicago Press. 15–44. https://doi.org/10.7208/9780226136790-003

Rosenthal, David M. 1986. Will and the Theory of Judgment. In Amélie Rorty (ed.) *Essays on Descartes' Meditations*. Berkeley, University of California Press. 405–434. https://doi.org/10.1525/9780520907836-020

Scotus, Johannes Duns 2016. *On Being and Cognition: Ordinatio 1.3*. Trans. John van den Bercken. New York, Fordham University Press. https://doi.org/10.2307/j.ctt1h1hx3p

Tamura, Ayumu. 2018. デカルト形而上学における「経験」概念に関する考察:「直観」および「知解」との対照において. 哲学・思想論叢. 36. 1–13. [In Japanese]

Tamura, Ayumu 2019. „*Experientia*" の系譜とデカルト:デカルト的「経験」概念の特性を究明するための予備的考察. 哲学・思想論叢. 37. 1–17. [In Japanese]

Attila Hangai

What is Rational Reconstruction in the History of Philosophy?

A Reply to Live Reconstructivists*

Christia Mercer's defence of the contextualist history of philosophy and her opposition to rational reconstruction in the *Journal of the History of Philosophy* (Mercer 2019) induced some direct reflection in the *Hungarian Philosophical Review* 66/1 (2022), titled *Reconstructivists Not Dead*. As the title suggests, the editors propose to defend reconstructivist methodology as the aim of the special issue:

> Our aim is twofold. First, to present methodological reflections on what exactly reconstructivist methodology consists in, how it is different from contextualism, and how it can provide new perspectives and insights not available for contextualists. Second, to demonstrate, with the help of case studies, that reconstructivist research can produce relevant and exciting new results. (Szalai and Tóth 2022. 6)

In what follows, I reflect on some arguments found in the issue and in Mercer's study: first, (I) on the purported distinction between reconstructivism and contextualism; second, (II) on Mercer's principle for making the distinction; third, (III) on what rational reconstruction amounts to.

I. RECONSTRUCTIVISM VS. CONTEXTUALISM

Let us start with the very distinction: reconstructivism versus contextualism. I think Mercer rightly identifies an important divide as a difference in purpose (Mercer 2019. 530, 533–539). Namely, whereas "contextualists" are engaged in a study to identify and explain views and arguments of past philosophers in their own historical context, that is, to explain *historical facts*, "reconstructivists" aim to gather *philosophical ideas* (positions and arguments) from historical texts, sometimes irrespective of whether the given philosopher did hold exactly such

* I am grateful to Márton Dornbach and Ágoston Guba for their comments that helped me improve the paper. All the remaining errors are my responsibility.

views;[1] and they do so to use these philosophical munitions in their contemporary philosophizing. Thus, the distinction is between philosophical and historical *interests*.[2]

Being a matter of aim and purpose, however, drawing the distinction in terms of methodology seems utterly misleading.[3] For one, it seems possible to aim at answering a philosophical problem through studying a past philosopher and their cultural context (philosophical aim and contextualist method), just as to aim at a correct identification of a philosopher's view through reconstructing her arguments preserved in a scattered way (historical aim and reconstruction as a method). For the former, there are numerous examples (quite a lot in Mercer's paper, which she categorizes as "contextualists"). For the latter, there are plenty of works, such as the whole area of Pre-Socratics and most of Hellenistic philosophy,[4] or Leibniz's metaphysics preserved in different writings and notebooks (on which see Blank 2022), to name just three. Even worse for Mercer's account, it is likely that any reasonable scholar reconstructing an old view (with philosophical or historical aim) will appeal to some context of the philosopher in question[5] (hence, they will also be "contextualists" for Mercer (even the notorious Della Rocca),[6] and so unsurprisingly she does not find any live "recon-

[1] As Schmaltz (2022. 27–28) notes, those who did not care at all about historical accuracy were not reconstructivists but philosophers dealing with strawman views.

[2] Many agree with the primacy of such a distinction, for example, Frede 2022; 1988; 1987; Normore 2016; 1990; Garber 2015; Vermeir 2013. 53–57; Lærke 2013; Hatfield 2005; Rorty, Schneewind, and Skinner 1984; Ayers 1978; Mandelbaum 1965. Also Rorty 1984, although he apparently confuses the terminology, on which see section III. Skinner (1969. 3–5) puts the distinction in terms of studying the text vs. studying the context. Passmore (1965) and Gracia (1992. 234–276) propose more detailed differentiations of aims.

[3] The terminology seems to originate from Rorty 1984, see section III.

[4] Cf. Frede 2022. 58, 85–86. While both Mercer and her critics restrict the scope of their reflections to early modernity, most of what they say applies to history of philosophy on a general, theoretical level. Thus, allowing that there will be differences in the specific methodology applied for material from different ages, and that the difference stems from the difference in the nature of the textual evidence, I aim to consider what is common in studies of philosophies in different periods.

[5] Cf. Rée 1978. 30; Frede 2022. 26, 30; Catana 2013. especially 133; Copenhaver 2020. However, Blank (2022. 72) denies using "contextualist methodology". This seems an adequate description of his paper, although he does appeal to many (even minor) works of Leibniz, which is sometimes cited as a contextualist feature (although, as it is often emphasized, the use of *all* relevant evidence is a minimal criterion of historical study, see, e.g., Normore 2016. 33, 45–46). Blank's reconstruction, however, is devoid of the feature which the methodological papers of the reviewed issue – Lenz 2022; Schmaltz 2022; and Tóth 2022 – emphasize: using concepts in the reconstruction of the view that are not available for the given author due to historical reasons. On this, see section II.

[6] Della Rocca is accused by Garber (2015) of doing mere rational reconstruction in Della Rocca 2008 instead of taking the job of the historian of philosophy seriously, that is, to identify the real historical Spinoza. Among many simplifications of Della Rocca, Garber points to the missing identification of Spinoza's motivations, especially his ethical and political concerns, while reducing his philosophy to pure rationalism following from the principle of sufficient reason. Della Rocca (2015) agrees that what he does is rational reconstruction. Moreover, he

structivist"). Thus, it would be better to label the main divide as between (say) "philosophical history of philosophy" and "historical history of philosophy".[7] This is not just a point of terminology, though, since this division and the methodological one are independent, as we have seen. Hence, it would be better to keep them distinct; so (I believe) we could avoid a lot of confusion.

Thus, this means that the distinction is not merely a matter of emphasis on different aspects of a "dialogical reading", as Martin Lenz suggests (Lenz 2022). He is right that we can learn much about our philosophical assumptions or the genealogy of our concepts by reading historical texts, as we inevitably apply our own conceptual framework to understand those texts. However, the very nature of his dialogical reading invites a division into two steps in the process. First, the reader aims to establish an interpretation of the historical view; then, she brings it into dialogue with our contemporary concepts or another text embodying our contemporary concepts. Granted that current considerations already influence the first step, it still relies on historical evidence and seems to remain faithful to the past author and their concepts.[8] So, Lenz's approach is best seen as a combination of historical reconstruction with a further step of philosophical reflection (similar to the idea expressed, e.g., by Rorty 1984. 49).[9]

At one point, however, Lenz insists on the difference between philosophical and historical *aims*, in addition to the methodological divide (Lenz 2022. 17). For an unstated reason, he takes his philosophical aim of learning about our assumptions as more prominent than any historical aim could be (although he only mentions a quite secondary one of extending the canon,[10] but not the explanation of historical facts). Accordingly, he construes his dialogical method as dependent on his *philosophical aim*. That is, Lenz fails to be in dialogue with Mercer, who writes about a methodology for *historical aims*.

---

acknowledges that his aims are philosophical: to criticize contemporary philosophy through Spinoza's philosophy. Thus, Della Rocca could have been a par excellence "reconstructivist" for Mercer (philosophical aim with reconstructivist method), but she misses the opportunity, and instead clings to Della Rocca's comment that he does believe to identify the real Spinoza. Mercer's taking Della Rocca to be a "contextualist" is all the more surprising, as she needs to stretch the meaning of "context" beyond plausibility to cover Della Rocca's principle of interpretation, namely, the principle of sufficient reason. I hope no argument is needed to show that this is not a contextual factor (in the sense used in the debate).

[7] This is the terminology of Frede 2022.

[8] This "dialogue" seems to be an application of our later concepts as a heuristic to search for similar or analogous conceptions in the past work. At the end, the past view must be identified through a thorough (historical) interpretation of the textual evidence, no matter what heuristics are applied in the process. As Gracia (1992. 154–155) points out, a real dialogue requires that the parties can respond to each other and all of them can (and is open to) change their views. Thus, strictly speaking, dialogue with the past is impossible.

[9] Cf. Beaney 2013. 253–255, who calls his own approach "dialectical reconstruction;" also see Frede 2022. 10, 26, 124–130; Passmore 1965. 12–13.

[10] This aspect is the focus of Klein 2022, so the placing of her paper in the journal issue is unclear. Hence, I do not discuss it here.

Tóth – who also confuses the aim of a study with the methodological question[11] – takes such a two-step approach (of historical reconstruction and philosophical reflection) as incoherent (Tóth 2022. 66). I see no incoherence here. His argument for the purported incoherence seems to depend on confusion. He suggests that contextual studies aim at identifying the actual view of a philosopher, which depends on the philosopher's *utterance*, whereas reconstructivists try to recover possible views, namely *propositions*. This curious wording often in the paper disguises that the contextual approach aims at historical truth without resorting to contemporary concepts,[12] while the reconstructivist aims at historical truth[13] by allowing concepts unavailable to the historical figure (on this question, see section II). In this specific case (of the alleged incoherence), Tóth is misled by his curious formulation: "the reconstructivist does not aim at reconstructing what historical agents took the utterance to mean, rather at discovering what proposition is expressed by the utterance." (Tóth 2022. 58). But what could be the intended meaning of an utterance, if not the propositions expressed by the utterance (with illocutionary force,[14] of course)? 'Proposition' is, after all, introduced into the philosophical vocabulary for the *meaning* of utterances of sentences and the like. That is, both contextual reconstruction and rational reconstruction will result in a set of propositions (and since we are discussing philosophical views, the propositions will have inferential relations among them). The difference between them is, perhaps, that while the contextual reconstruction aims to identify the set of propositions most plausibly attributable to the past author

[11] Let me pick out just one of his rhetorical arguments: appeal to the (putative) absurdity of thinking that we understand Aristotle better than Aquinas simply due to our better grasp of the context (Tóth 2022. 62). This completely misses the mark. First, Aquinas did not engage in history of philosophy: he pursues philosophy, just as any commentator before him and many after him. Thus, it is quite likely that historians of philosophy indeed have a better grasp of the historical Aristotle than Aquinas, allowing that Aquinas' philosophical interpretation or reconstruction is quite ingenious. Second, reserving such an authority to the Great Aquinas is quite at odds with Tóth's general audacity for claiming that rational reconstructivists often better understand past philosophical views than the past philosophers themselves.

[12] In this paper, I am not focusing on issues with "contextualism" – which I discuss in Hangai (2023) –, so I only mention cursorily here that Tóth's circumscription of contextualism as a search for identifying the *particular intention* the past philosopher had (Tóth 2022. 55–57) (hence, the meaning of the text) is, in a sense, a distortion. Historical facts need not be about the *psychology* of an agent or an author. Indeed, psychological explanations should be avoided in history of philosophy as much as possible, while explanation in terms of philosophical reasons is to be preferred. See Frede 2022; 1988; 1987; also Skinner 2002b, Gracia 1992. 229–231, *passim*. That is, an internal history of philosophy (even if called "contextualist") need not (and should not) privilege the author's actual intention or interpretation, but should take into account interpretations of contemporaries of the author, those that contributed to the same debate as the author. Cf. Skinner 2002c. 77–78; 2002a. 110–111. Also see the accounts of Passmore 1965; Mandelbaum 1965; Normore 1990; 2016; Hatfield 2005. 103–110; Lærke 2013; Vermeir 2013.

[13] At some points, he makes the distinction in terms of purpose (reconstructivists having philosophical interest), for example, Tóth 2022. 64.

[14] See Skinner 2002a; 2002c. 82; 1969. 45–47.

(identifying the historical truth), the rational reconstruction does not do so.[15] So, it is unclear why it would be incoherent to assess, once determined, such a set of propositions according to current philosophical standards.

II. MERCER'S PRINCIPLE

Let us turn to how Mercer conceives of the main distinction above. She provides a minimalist criterion to count a given approach as historical (or contextualist): the aim to "getting things right", that is, to be historically accurate. Influenced by Skinner (1969), she explicates the rule negatively in terms of the historical author's *psychology*: "historians of philosophy should not attribute claims or ideas to historical figures without concern for whether or not they are ones the figures would recognize as their own" (Mercer 2019. 530). Since there is no independent criterion of whether the psychological criterion holds (we cannot ask dead philosophers whether they would adopt a view or recognize it as their own) but the historical (mainly textual) evidence, we can reduce Mercer's rule to conformity with ordinary historical evidence.[16]

Schmaltz criticizes Mercer's principle as not allowing the use of concepts unavailable to the past philosophers in the reconstruction of their views (Schmaltz 2022). In practice, such would be contemporary concepts (or at least later than the historical author), like referential opacity, as Quine conceived it. Schmaltz says the possibility that an interpretation appealing to such a concept C is correct implies that what the past author A *meant* is to be understood in terms of concept C. Since concept C is supposed to be unavailable to author A, A could not recognize the reconstructed view as their own since it involves concept C (and concepts unavailable to one cannot be recognized by one). Hence, Mercer's principle would render the interpretation incorrect; as Schmaltz puts it, it would "violate Mercer's GTRC,[17] as much as Skinner's principle"[18] (Schmaltz 2022. 30). While I agree that using such concepts in interpreting past authors is not straightforward, I think doing so does not necessarily violate Skinner's and Mercer's principle. As I will suggest interpreting the past author's claim in terms of such a concept C involves attributing the availability of concept C to the author rather than appealing to a concept unavailable to them. Before turning to the details, let us review Schmaltz's arguments for his claim.

[15] Curiously, Tóth believes that rational reconstruction does consider all the historical data and context, so he remains closer to contextualism than he admits. On my remarks on this, see section II.

[16] This point is already made clear by Skinner himself, see Skinner 1969. 28–29, 49. On this point, see also Schliesser 2019.

[17] Mercer abbreviates her "getting things right constraint" as "GTRC".

[18] For the original explication, see Skinner 1969. 28.

In support of this view, he appeals to the ordinary case that sometimes an interpreter might hit the mark better than a speaker or, in general, an agent; that is, the interpreter understands better what the agent means than the agent itself. The reason cited (from Kant) is that the agent "may not have determined his concept sufficiently" (Kant 1998. A314/B370). Before turning to the ordinary phenomenon, let me comment on the appeal to Kant's authority. First, if Kant is right in his diagnosis of our better understanding than the author, the concept will appear to be *available* to the author. For according to Kant, the author did not have an accurate understanding because they determined the concept insufficiently. But insufficient determination is already some determination, so it seems clear that according to Kant, the author already has the concept in some (though insufficiently determined) way. So, the concept is available to the author, at least in principle.[19] Second, Kant is explicit in the sentence preceding the cited one that he does not mean his interpretation (perhaps appealing to concepts insufficiently determined by their authors) to be a historical interpretation, or as Kant writes, "I do not wish to go into any literary investigation here, in order to make out the sense which the sublime philosopher combined with his word." (*ibid.*) So, the citation of Kant should rather imply (if we want to draw any conclusion from Kant's authority) that when we better understand some author than the author themselves, even then, the author has access to all the concepts in terms of which we understand them, although in an insufficient way. But, again, such a case should not be taken as a historical understanding anyway.[20]

The ordinary phenomenon – we understand better what a speaker means than the speaker – is still pressing, though. But consider the case more fully: it typically continues like this. Once we express our (better) interpretation to the speaker, the speaker *acknowledges* that it is indeed a better understanding of what they meant. This acknowledgement, then, is to be counted as a criterion of what the meaning actually was.[21] In extending this case to writings and past authors, however, we lose this kind of criterion, as they cannot make acknowledgement. However, without such a criterion, it is unclear how we could know if our understanding of what the author meant is better than that of the author. One obvious criterion is the historical evidence, or the "context" if you like. But even though Schmaltz admits the importance of historical evidence (Schmaltz 2022. 6–7), he seems to deny that historical evidence would be decisive in such a case.

Return to Schmaltz's assumption that an interpretation appealing to a concept unknown to a past author might be correct. Let me construe the situa-

[19] Cf. Kant (1998. B9–11, A8) on how analysis of concepts leads to better understanding of them without adding anything to them. Cf. Passmore 1965. 32.

[20] For a similar reading of Kant's better understanding the author, see Dornbach 2016. 90–97.

[21] Compare Grice (1989, especially Part I: Logic and Conversation). Cf. Rorty 1984. 53–54.

tion differently, which disarms the case against Mercer's constraint. What if the interpretation using concept C turns out to be correct? First, as I suggested above, this should imply that the historical evidence is closely scrutinized, and the interpretation is determined to be adequate to it and more adequate than competing interpretations. But what about concept C? I surmise that in such a case, the adequate interpretation should demonstrate – with supporting textual evidence – that author A indeed had access to concept C (at least, perhaps, in an insufficient way, as Kant suggested). Thus, it would turn out that the correct interpretation will not involve concepts unavailable to the author in question; rather, it would involve attributing the availability of concept C to the author that previously was thought to be unavailable to them.[22] In terms of Mercer's principle, we could say that once A was faced with the interpretation of A's view that involves concept C, A could recognize it as her own, which would comprise A's recognition (learning) of concept C.[23]

It is indeed questionable whether we should allow using such concepts to interpret past authors.[24] Schmaltz himself proposes to be cautious, perhaps leaving it for last resort (Schmaltz 2022. 6–7). With the caution, Tóth (2022. 61) agrees while making the possibility of using concepts unavailable to a past author in historical interpretations the mark of his own "reconstructivism".[25] His caution is, however, telling. He says we should appeal to unavailable concepts only "if

[22] This is how Szalai (2022) proceeds: attributing the distinction between narrow and broad content to Descartes based primarily on textual evidence where Descartes does use such a distinction, so that she is able to account for certain claims of Descartes that otherwise would lack explanation. While she admits the "apparent anachronism", the lack of a label for and systematic treatment of the distinction, and the shifts between Descartes' appeal to narrow and broad content within short passages (*ibid.* 102), we might take these as signs that Descartes did have the concept in an unclear way (since he did use the distinction), rather than that he did not have it at all. A note on the purported anachronism. Since many commentators of Aristotle interpret the distinction of regarding a *phantasma* in itself and as a copy (or image, *eikōn*) of something else at *de Memoria* 450b20–27 exactly as the distinction of narrow and broad content *à la* Szalai (this seems to be the "orthodox reading" in 20th century scholarship, see references in Caston 2021. note 45 and 47), it seems unlikely that this distinction was completely unfamiliar in the 17th century. However, Caston argues against this reading of Aristotle (Caston 2021. 177–190). Again, the distinction between narrow and broad content might be similar to the case of the concept of subjectivity, which Kaukua and Lähteenmäki 2010 take as a non-textual standard of interpretation that we should assume to be available anytime.

[23] For an approach similar to what I propose in the text, suggested by James Conant, building indeed on Kant's "better understanding", see Forsberg and Conant 2013. 155–159. I owe this reference to Márton Dornbach.

[24] See especially Skinner 2002c. 59–60, 77–78; also Skinner 1969. 7–9, 22–24, 28–29, where the relationship between views of politics on the one hand, and contemporary political practice (or later significance of the views in question) on the other, makes Skinner's case decisive.

[25] Tóth's proposal is probably similar to Della Rocca's holistic approach, according to which the meaning of a historical text depends on contemporary conceptual frameworks, just as everything else; and so history of philosophy is not different from philosophy after all (Della Rocca 2020. 194–195).

the assumption that the historical author expressed that view explains more features of the historical author's utterance than rival interpretations" that appeal to the historical context. That is, the historical context can never be ignored and sets a constraint on what interpretation is acceptable. Thus, Tóth seems unsuccessful in distancing himself from the "contextualist" methodology. Again, it seems unlikely that an interpretation that appeals to a concept C unavailable to philosopher A can explain "more features" than one that appeals only to concepts certainly available to A. An explanation of the former kind leaves at least one "feature" unexplained: the use of concept C.

Lenz, touching upon the issue, goes further. He takes it as a historicist principle that concepts unavailable to an author should not be used in the interpretation of the author's text (Lenz 2022. 12) while submitting that, to gain a better understanding of the past text, it is inevitable to apply such contemporary concepts (or concepts developed significantly later than the target author). His reasons are, however, blatantly flawed (Lenz 2022. 13). First, he suggests that "writing for future generations," which seems a relatively common practice of philosophers,[26] implies that the interpretation of a text that ignores the future reception of the text (this appears to be the idea) is poor. I do not see how this is supposed to follow, and Schliesser's idea of philosophical prophecy (Schliesser 2013) does not help. For a text's meaning differs from how it is understood later (the significance and reception).[27] Second, Lenz suggests that we cannot understand a sentence without taking it to be either true or false. It is true that I cannot fail to take the "that"-clause in the previous sentence to be false *once I have understood it* – but this is completely different from saying that I understand its meaning *because I take it to be false*. Lenz construes the dependency the wrong way. In any case, since the understanding of the sentence does not (and cannot) depend on taking it to be true or false, I can fully understand any sentence without assigning a truth value to it.[28]

[26] Cf. Frede 2022. 28.

[27] Cf. Skinner 1969. 23. As Normore (2016. 43) has put it, "a misreading of a text can be as historically important as a correct reading!" Note that this is consistent with the legitimacy of questions of reception in the history of philosophy, although such questions are different from the question what a given philosopher meant; and, as Hatfield (2005. 106–109) submits, come later than the more basic question of identifying what the author meant in its own context.

[28] This is lucidly shown by Skinner (2002b. especially 29–30). Also see Glock 2008. 889–892.

III. RATIONAL RECONSTRUCTION: WHAT IS IT?

While the papers discussed (and especially the issue editors) aim to defend rational reconstruction as a valid method in the history of philosophy (or in philosophy, for that matter), they use the term mainly in the sense Rorty used it, as I will show it shortly, and give little or no reflection on what this method amounts to (one exception is Tóth 2022). This is all the more problematic, for some confusions detected above could be avoided with reflection. To get a clearer view of what "rational reconstruction" is, let me close with a brief recapitulation of the history of the concept based on Michael Beaney's succinct account and Imre Lakatos' view of rational reconstruction in the history of science. With these in hand, we can find a place for rational reconstruction in history of philosophy. But first, let us see Rorty's account.

Rorty describes rational reconstruction as applying our concepts to past authors to engage in a philosophical discussion with them (Rorty 1984. 49–56). Thus, he emphasized the distorting, anachronistic tendency of rational reconstruction that forces *contemporary concepts* into past views so that they appear as alternative answers to contemporary philosophical problems. Yet, he also proposes to do historical reconstruction, involving an appeal to the cultural context and, most importantly, Skinner's principle as a constraint of accuracy. Simply put, Rorty makes the distinction in terms of Skinner's principle (just like Mercer, as we have seen in section I): historical reconstruction adopts it – rational reconstruction does not (ibid. 54). He prefers doing both separately as the two-step approach mentioned above (ibid. 49), where reflection on our concepts follows historical reconstruction. However, he argues for the inseparability of the two methods, pointing to the indispensability of our concepts to understanding other's acts and language (ibid. 51, note 1).[29] Whichever is Rorty's preferred view, most of his considerations point to the former, and he does not sufficiently specify the latter. Thus, it seems the debate related to Mercer's paper – that emphasizes the use of concepts unavailable to a past author in the interpretation – remains in the confines of Rorty's framework and so is misled by Rorty's terminology.

Turning to the history of the method, Beaney (2013. 233–236) detects two sources of rational reconstruction: Neo-Kantianism and Logicism. Neo-Kantians emphasized the distinction between discovery and justification. While discovery (or the generation of our beliefs) is explained mainly in psychological terms, and so it has spurious value, justification (and validity of knowledge) follows strict rules, so the norms guiding justification can be studied scientifically. The method for the latter is called "critical" or "reconstructive". Again, Frege's reduction of arithmetic to logic can be seen as a reconstruction. What he does is,

[29] The discrepancy is also noted by Beaney (2013. 247).

first, clear the ground from mistaken views and identify aspects of the concept under investigation (the concept "number"), which is a sort of historical investigation. Second, he construes (reconceives) arithmetic in terms of concepts and logical relations.

Following these traditions, "rational reconstruction" becomes an explicit philosophical method for Carnap and subsequently for Popper, Reichenbach, and others (Beaney 2013. 237–242). The authors apply rational reconstruction to various problems, so the method is not uniform for them, but something of the following emerges. Rational reconstruction has to do with *explication* (or redefinition) of our old concepts using clearer (or more basic) concepts, paying particular attention to identifying the *logical relations* between concepts, and aiming to arrive at a more or less *systematic* and *coherent structure* of concepts. Thus, the method is essentially *normative*: it prescribes coherence and systematicity in logical relations. Moreover, it is used primarily in the logic of justification, where *time* (hence history) is not a factor.

Nevertheless, rational reconstruction has been applied in the history of philosophy (as we have seen) and in the history of science. The idea can be fruitfully elaborated using the influential paper of Imre Lakatos (1971). Lakatos argues that an epistemological theory of scientific discovery – inductivism, conventionalism (e.g., Duhem), falsificationism (like Popper), and his methodology of research programs – entails a historical narrative of scientific development. The methodology provides a normative rule of rationality to demarcate science and a criterion for what counts as an internal rational history of science (a rational reconstruction) and what remains as an external irrational factor (e.g., psychological, sociological context). Even though the history of science needs external history as well, internal history is the primary in two ways. First, the internal history of science is autonomous (since this is the rational part of science); hence understanding the logic of science does not require external history. Second, the questions in external history depend on internal history.

Lakatos argues for his methodology by comparing the adequacy of the histories entailed by the different rationality norms of science. That is, he compares the rational reconstructions of the history of science according to each demarcation criterion: how much each renders the actual history of science rational. He prefers the reconstruction that fits the actual history (historic events) better (especially ibid. 117–118). The role remaining for external history is telling: "either provides non-rational explanation of the speed, locality, selectiveness etc. of historic events as *interpreted* in terms of internal history; or, when history differs from its rational reconstruction, it provides an empirical explanation of why it differs" (ibid. 105–106). Thus, importantly, the history of science for Lakatos should explain all the past of science and explain it internally and rationally as much as possible; where it is not possible to provide a rational explanation, an empirical explanation of external history should be supplied. This is reminis-

cent of Michael Frede's approach to the history of philosophy (Frede 2022), which prefers internal history – as much a rational endeavour as possible – over external history (appealing to all sorts of cultural contexts). For him, contextual factors are allowed to *enter into the internal history* to provide empirical explanations when the purely rational explanation is impossible[30] (for Frede, the external history contains all sorts of inter-relations with other scientific disciplines, theology, sociological or cultural context, and psychological motivations). Thus, let me apply Lakatos' view on the history of philosophy by paraphrasing his paraphrase of Kant (Lakatos 1971. 91): "Rational reconstruction without actual history of philosophy is empty; actual history of philosophy without rational reconstruction is blind."

The normative aspect of rational reconstruction and the fact that it can easily lead to distorting selection is apparent in examples like Russell, who aims at identifying generic types of philosophy rather than particular historical views and forces a logically coherent system on the opinions of philosophers (Beaney 2013. 248–252). Another aspect of normativity is the connection to criticism. Like Russell, the so-called "reconstructivists" do not stop with the systematic re-description of the views; they *criticize* anything that does not fit into or conflict with the system. Rational reconstructions "have an important role in making us aware of the logical relations between the views a philosopher holds and facilitating assessment of the validity and soundness of their arguments" (Beaney 2013. 253).

Apparently, Rorty picks up this latter aspect while leaving the positive role of the normativity of rational reconstruction untouched, especially in demarcating internal from external history. Instead, for similar purposes, Rorty introduces two other (larger scale) "genres" of history of philosophy besides historical and rational reconstruction: *Geistesgeschichte*, for guiding larger-scale historical narratives and giving a rationality principle to the history of philosophy (Rorty 1984. 56–61); and *intellectual history* for Geistesgeschichte to remain honest (ibid. 67–74). A simpler, more economical (and preferable) alternative could be construed along the lines of Lakatos' account. It would assign a normative role to rational reconstruction in determining the internal history of philosophy while the actual historical facts would be explained thoroughly (through historical reconstruction, if you will), primarily as rational philosophical acts, secondarily as effects

[30] I make the comparison only with regard to this aspect. Most importantly, Frede by no means suggests (or gives a hint) that the history of philosophy should be reconstructed in accordance with a uniform norm of philosophical rationality. But the status of non-rational, contextual factors as empirical explanations of historical facts unexplained rationally is quite important for him. See especially Frede 2022. 53–54, 84, 100–101; 1987. xi–xviii; 1988. 669–672. For a similar approach (influenced by Frede), see Normore 1990. 221–226 and 2016. 38–42.

of contextual factors of the time. As suggested above, Frede's account can be construed along these lines.[31]

Let me close by approvingly citing Beaney's reflection on Quine's critique of rational reconstruction (à la Rorty et al.) devoid of historical accuracy (Beaney 2013. 244):

> As I see it, Quine raises a dilemma here for any project of rational reconstruction. Either rational reconstruction aims to provide translational equivalents, or it does not. If it does, then all well and good, but no attempts have yet been successful. If it does not, then there will always be something to explain, in which case appeal will need to be made to actual history (or psychological genesis). But if such appeal is needed, then why not seek to explain the actual history in the first place?

REFERENCES

Ayers, Michael 1978. Analytical Philosophy and the History of Philosophy. In Jonathan Rée – Michael Ayers – Adam Westoby (Eds.) *Philosophy and its Past*. Hassocks, Harvester Press. 41–66.

Beaney, Michael 2013. Analytic Philosophy and History of Philosophy: The Development of the Idea of Rational Reconstruction. In Erich H. Reck (Ed.) *The Historical Turn in Analytic Philosophy*. London, Palgrave Macmillan. 231–260. https://doi.org/10.1057/9781137304872.0017

Blank, Andreas 2022. On Reconstructing Leibniz's Metaphysics. *Hungarian Philosophical Review*. 66/1. 69–89.

Caston, Victor 2021. Aristotle and the Cartesian Theatre. In Pavel Gregoric – Jakob Leth Fink (Eds.) *Encounters with Aristotelian Philosophy of Mind*. New York, Routledge. 169–220. https://doi.org/10.4324/9781003008484-11

Catana, Leo 2013. Philosophical Problems in the History of Philosophy: What Are They? In Mogens Laerke – Justin E. H. Smith – Eric Schliesser (Eds.) *Philosophy and Its History*. Oxford, Oxford University Press. 115–133. https://doi.org/10.1093/acprof:oso/9780199857142.003.0007

Copenhaver, Brian 2020. A Normative Historiography of Philosophy: Room for Internalism and Externalism. *British Journal for the History of Philosophy*. 28/1. 177–199. https://doi.org/10.1080/09608788.2019.1608903

Della Rocca, Michael 2008. *Spinoza*. London, Routledge. https://doi.org/10.4324/9780203894583

[31] One might suggest that retrospective histories of philosophy that proceed from a given philosophy and trace its genesis back in the past, like Hegel's history of philosophy is more akin to the structure of Lakatos' account, insofar as such histories apply a given view of philosophy (and so philosophical rationality) to the entire history of philosophy. This is true, but just as Lakatos demonstrates that the theories alternative to his (inductivism, conventionalism, and falsificationism) lead to inadequate historical accounts, inadequacy could be shown for any history of philosophy that applies a uniform norm of rationality throughout the history of philosophy. It seems clear that what philosophy was taken to be in the past (say by Plato) is quite different than what philosophy is taken to be nowadays.

Della Rocca, Michael 2015. Interpreting Spinoza: The Real Is The Rational. *Journal of the History of Philosophy*. 53/3. 523–535. https://doi.org/10.1353/hph.2015.0049

Della Rocca, Michael 2020. *The Parmenidean Ascent*. New York, Oxford University Press. https://doi.org/10.1093/oso/9780197510940.001.0001

Dornbach, Márton 2016. *Receptive Spirit: German Idealism and the Dynamics of Cultural Transmission*. New York, Fordham University Press. DOI: https://doi.org/10.1515/9780823268313208

Forsberg, Niklas – James Conant 2013. Interview. From Positivist Rabbi to Resolute Reader: James Conant in Conversation with Niklas Forsberg, Part 1. *Nordic Wittgenstein Review*. 2/1. 131–160. https://doi.org/10.1515/nwr.2013.2.1.131

Frede, Michael 1987. Introduction: The Study of Ancient Philosophy. In his *Essays in Ancient Philosophy*. Oxford, Oxford University Press. ix–xxvii.

Frede, Michael 1988. The History of Philosophy as a Discipline. *The Journal of Philosophy*. 85/11. 666–672. https://doi.org/10.5840/jphil1988851114

Frede, Michael 2022. *The Historiography of Philosophy* (ed. Katerina Ierodiakonou). New York, Oxford University Press. https://doi.org/10.1093/oso/9780198840725.001.0001

Garber, Daniel 2015. Superheroes in the History of Philosophy: Spinoza, Super-Rationalist. *Journal of the History of Philosophy*. 53/3. 507–521. https://doi.org/10.1353/hph.2015.0045

Glock, Hans-Johann 2008. Analytic Philosophy and History: A Mismatch? *Mind*. 117/468. 867–897. https://doi.org/10.1093/mind/fzn055

Gracia, Jorge J. E. 1992. *Philosophy and Its History: Issues in Philosophical Historiography*. Albany/NY, State University of New York Press.

Grice, Herbert Paul 1989. *Studies in the Way of Words*. Cambridge/MA, Harvard University Press.

Hangai, Attila 2023. Internal History of Philosophy. *Filozofia*. 78/10. 848–864. https://doi.org/10.31577/filozofia.2023.78.10.4.

Hatfield, Gary 2005. The History of Philosophy as Philosophy. In Tom Sorell – John Rogers (Eds.) *Analytic Philosophy and History of Philosophy*. Oxford, Oxford University Press. 89–128. https://doi.org/10.1093/oso/9780199278992.003.0006

Kant, Immanuel 1998. *Critique of Pure Reason*. Trans. Paul Guyer – Allen W. Wood. Cambridge, Cambridge University Press.

Kaukua, Jari – Lähteenmäki, Vili 2010. Subjectivity as a Non-Textual Standard of Interpretation in the History of Philosophical Psychology. *History and Theory*. 49/1. 21–37. https://doi.org/10.1111/j.1468-2303.2010.00526.x

Klein, Julie R. 2022. The Past and Future of the Present. *Hungarian Philosophical Review*. 66/1. 35–49.

Lærke, Mogens 2013. The Anthropological Analogy and the Constitution of Historical Perspectivism. In Mogens Laerke – Justin E. H. Smith – Eric Schliesser (Eds.) *Philosophy and Its History*. Oxford, Oxford University Press. 7–29. https://doi.org/10.1093/acprof:oso/9780199857142.003.0002

Lakatos, Imre 1971. History of Science and its Rational Reconstructions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1970. 91–136. https://doi.org/10.1086/psaprocbienmeetp.1970.495757

Lenz, Martin 2022. Did Descartes Read Wittgenstein? A Dialogical Approach to Reading. *Hungarian Philosophical Review*. 66/1. 9–23.

Mandelbaum, Maurice 1965. The History of Ideas, Intellectual History, and the History of Philosophy. *History and Theory*. 5/5. 33–66. https://doi.org/10.2307/2504118

Mercer, Christia 2019. The Contextualist Revolution in Early Modern Philosophy. *Journal of the History of Philosophy*. 57/3. 529–548. https://doi.org/10.1353/hph.2019.0057

Normore, Calvin 2016. The Methodology of the History of Philosophy. In Herman Cappelen – Tamar Szabó Gendler – John Hawthorne (Eds.) *The Oxford Handbook of Philosoph-*

ical Methodology. Oxford, Oxford University Press. 27–48. https://doi.org/10.1093/oxfordhb/9780199668779.013.17

Normore, Calvin 1990. Doxology and the History of Philosophy. *Canadian Journal of Philosophy, Supplementary Volume*. 16. 203–226. https://doi.org/10.1080/00455091.1990.10717226

Passmore, John 1965. The Idea of a History of Philosophy. *History and Theory*. 5/5. 1–32. https://doi.org/10.2307/2504117

Rée, Jonathan 1978. Philosophy and the History of Philosophy. In Jonathan Rée – Michael Ayers – Adam Westoby (Eds.) *Philosophy and its Past*. Hassocks, Harvester Press. 1–39.

Rorty, Richard 1984. The Historiography of Philosophy: Four Genres. In Jerome B. Schneewind – Quentin Skinner – Richard Rorty (Eds.) *Philosophy in History: Essays in the Historiography of Philosophy*. Cambridge, Cambridge University Press. 49–76. Rorty, Richard 1984. https://doi.org/10.1017/cbo9780511625534.006

Rorty, Richard – Jerome B. Schneewind – Quentin Skinner 1984. Introduction. In their *Philosophy in History*. Cambridge, Cambridge University Press. 1–14. https://doi.org/10.1017/CBO9780511625534.003

Schliesser, Eric 2013. Philosophic Prophecy. In Mogens Laerke – Justin E. H. Smith – Eric Schliesser (Eds.) *Philosophy and Its History*. Oxford, Oxford University Press. 209–235. https://doi.org/10.1093/acprof:oso/9780199857142.003.0011

Schliesser, Eric 2019. On Getting Things Right (Constraints) in the History of Philosophy. *Digressions and Impressions* (blog). https://digressionsnimpressions.typepad.com/digressionsimpressions/2019/07/on-getting-things-right-constraints-in-the-history-of-philosophy.html.

Schmaltz, Tad M. 2022. Getting Things Right in the History of Philosophy. *Hungarian Philosophical Review*. 66/1. 25–33.

Skinner, Quentin 1969. Meaning and Understanding in the History of Ideas. *History and Theory*. 8/1. 3–53. https://doi.org/10.2307/2504188

Skinner, Quentin 2002a. Interpretation and the Understanding of Speech Acts. In his *Visions of Politics. Volume 1. Regarding Method*. Cambridge, Cambridge University Press. 103–127.

Skinner, Quentin 2002b. Interpretation, Rationality and Truth. In his *Visions of Politics. Volume 1. Regarding Method*. Cambridge, Cambridge University Press. 27–56. DOI: https://doi.org/10.1017/CBO9780511790812.009.

Skinner, Quentin 2002c. Meaning and Understanding in the History of Ideas. In his *Visions of Politics. Volume 1. Regarding Method*. Cambridge, Cambridge University Press. 57–89. DOI: https://doi.org/10.1017/CBO9780511790812.009.

Szalai, Judit 2022. Transparency and Broad Content in Descartes. *Hungarian Philosophical Review*. 66/1. 91–108.

Szalai, Judit – Tóth, Olivér István 2022. Reconstructivism Not Dead: Introduction. *Hungarian Philosophical Review*. 66/1. 5–8.

Tóth, Olivér István 2022. A Defense of Reconstructivism. *Hungarian Philosophical Review*. 66/1. 51–68.

Vermeir, Koen 2013. Philosophy and Genealogy: Ways of Writing History of Philosophy. In Mogens Laerke – Justin E. H. Smith – Eric Schliesser (Eds.) *Philosophy and Its History*. Oxford, Oxford University Press. 50–70. https://doi.org/10.1093/acprof:oso/9780199857142.003.0004

Contributors

Erik Åkerlund holds a PhD in theoretical philosophy with an emphasis on history of philosophy from Uppsala University (2011). He is Lecturer in Philosophy, and Director of Studies, at The Newman Institute, a Jesuit run university college in Uppsala, Sweden. Recently, Erik has been involved in the research project *The Mechanization of Philosophy 1300–1700*, funded by the Swedish Research Council (grant 2019–02777), based at Stockholm University, Sweden.
E-mail address: erik.akerlund@newman.se

László Bernáth is a Research Fellow at the Institute of Philosophy of the HUN-REN Research Centre for the Humanities. Areas of specialisation: metaphysics, epistemology, ethics.
E-mail address: Bernath.Laszlo@abtk.hu

Attila Hangai is a postdoctoral research fellow at the Institute of Philosophy of the HUN-REN Research Centre for the Humanities. He specializes in Ancient philosophy, especially Aristotle and the late antique Aristotelian and Platonist tradition, especially on issues in philosophical psychology and epistemology.
E-mail address: Hangai.Attila@abtk.hu

Ferenc Huoranszki is professor of philosophy at the Philosophy Department of Central European University. His interests include metaphysics and the philosophy of action, particularly the questions of free will, laws of nature, modality, and the history of 18th and early 19th century philosophy.
E-mail address: Huoransz@ceu.edu

Gergely Kertész is a philosopher of science, focused mainly on the philosophy of the life sciences, including psychology. He holds a philosophy degree from ELTE (2006), where he also studied informatics and literature. Following further HPS studies and fellowships at the Philosophy and History of Science Department, Budapest University of Technology and Economics (BME), he earned a PhD at the Philosophy Department of the University of Durham in the UK (2019). He taught at ELTE, BME, Durham, and various special colleges in Hungary. From 2019 to 2023, he worked as a researcher at ELKH-BTK (formerly MTA-BTK) in two 'Lendület' projects, 'Morals

and Science' and 'Values and Science'. From 2012, alongside his academic work, he works in the business sector. In 2023, he left academia for the business sector.
E-mail address: kerteszgergely@freemail.hu

Gyula Klima is professor of philosophy at Fordham University in New York, Doctor of the Hungarian Academy of Sciences, the founding director of two international scientific societies, and the author, translator, and editor of numerous books and studies, primarily dealing with the comparison of scholastic and contemporary logical and metaphysical theories.
E-mail address: klima@fordham.edu

Dániel Kodaj is assistant professor at the Institute of Philosophy at Eötvös Loránd University (ELTE). He works mainly on metaphysics, the philosophy of religion, and logic.
E-mail address: kodaj.daniel@btk.elte.hu. Homepage: dkodaj.net

Mohsen Moghri is a UKRI-MSCA Postdoctoral Fellow at the Department of Philosophy, University of Birmingham, and Visiting Researcher at the University of Religions and Denominations. His research primarily focuses on the metaphysical explanation of existence and related inquiries concerning axiological explanation, metaphysical grounding, and religious naturalism.
E-mail address: m.moghri@bham.ac.uk

Michael Ruse is the former Lucyle T. Werkmeister Professor and Director of the HPS Program at Florida State University. He works on the philosophy of biology, ethics, and the history and philosophy of science. His latest book is *A Philosopher Looks at Human Beings*.
E-mail address: mruse@fsu.edu

Ayumu Tamura is assistant professor at National Institute of Technology, Ibaraki College (JP). He works mainly in early modern philosophy, particularly Descartes. His recent articles are: Trace of Stoic Logic in Descartes: Stoic Axiōma and Descartes's Pronuntiatum in the Second Meditation. *The Seventeenth Century* (2023); Foucault–Derrida Debate on Madness Revisited. *Tetsugaku* (2023).
Email: atamura@gm.ibaraki-ct.ac.jp

Summaries

Erik Åkerlund

Models of Finality: Aristotle, Buridan, and Averroes

Final causation was famously problematized in early modern philosophy by Descartes, Spinoza, and others. However, the understanding of final causation had never been monolithic, and early modern philosophers can rather be understood as carrying on earlier philosophical traditions, especially from the transformations of philosophical debates in the late Middle Ages. However, due in large part to the literary style of Scholastic philosophy, it takes some work to decipher where the actual dividing lines lie in these debates. In this article, a heuristic scheme is offered, against the background of which the debates on finality in the late Middle Ages can hopefully be better accounted for. Based on a division between intentional and non-intentional conceptions of finality, on the one hand, and a division between what is here called Dynamic and Boolean metaphysics, respectively, on the other, three earlier thinkers are offered as paradigmatic examples of different models of finality: Aristotle (non-intentional and Dynamic), Buridan (intentional and Boolean), and Averroes (intentional and Dynamic).
Keywords: Aristotle, Averroes, Buridan, teleology

László Bernáth

The Aporia of Categorical Obligations and an Augustinian Teleological Way Out of It

There is a vast literature on the problem of how categorical obligations and reasons are possible. In this paper, I attempt to reconstruct both parts of this problem as clearly and as briefly as possible by spelling out a (not too extreme) view about relations between hypothetical reasons/obligations and categorical reasons/obligations. After that, I give a simple argument – based on one of Alasdair MacIntyre's examples – for the impossibility of categorical reasons. In the third section, I argue that it is not as big a problem as most suppose because (quasi-)categorical obligations are possible even without categorical reasons. In the last section, I propose an Augustinian view of motivational states and morality that shows the way out of the aporia by providing a plausible theory of (quasi-) categorical obligations (with a non-negligible ontological price tag).
Keywords: Augustinian anthropology, categorical obligation, categorical reason, teleology, value

Attila Hangai

What is Rational Reconstruction in the History of Philosophy? A Reply to Live Reconstructivists

Hungarian Philosophical Review 66/1 (2022) proceeds from discussing Christia Mercer's paper *The Contextualist Revolution in Early Modern Philosophy*. In this paper, I reflect on some arguments found in the journal issue and Mercer's study: first, on the purported distinction between reconstructivism and contextualism; second, on Mercer's principle for making the distinction; third, on what rational reconstruction amounts to.
Keywords: contextualism, history of philosophy, Mercer, reconstructivism

Ferenc Huoranszki

Intentional Actions and Final Causes

What distinguishes agents' intentional actions from those episodes in their life that merely happen to them? This paper argues that the intentionality of agents' actions is an irreducibly teleological phenomenon. An intentional action is a process that occurs for the sake of an end that we ascribe to the agent who performs it. This intrinsic teleological structure is a precondition, rather than a causal consequence, of human agents' capacity to mentally represent and consciously initiate their actions. Hence teleology is an intrinsic, and not a derivative, feature of the process in which agents who act participate. More specifically, the paper argues for two major claims. First, whenever agents' actions are intentional there must be a sense in which what they do is not a mere accident. The paper shows that the sense in which intentional actions are not accidents can only be explained with reference to the actions' final, rather than their efficient, causes. Second, it argues that it is the intrinsic teleological structure of actions that best explains the sense in which agents always try to do what they do intentionally.
Keywords: action theory, intentions, efficient causes, final causes, teleology

Gergely Kertész

On the Status of Teleological Discourse: A Confusing Fiction or a Description of Reality?

In modern philosophy and science teleological descriptions of nature got discredited and abolished from the mainstream worldview. With the advent of new theories of organisms and self-maintaining systems more generally a rethinking of the received view is in order and is already under way. This paper aims at assessing different possible interpretations of the status of teleological descriptions of organic, animate nature, considering the virtues and challenges of a realist, but physicalist/reductionist approach, comparing it at certain points to fictionalist and eliminativist attitudes. The aim is to establish that it is a live option, it is rational to think that teleology is a real, not purely projected property of some systems in nature. By "real" I don't mean that it is an ontologically fundamental property of physical simples. I aim to show that it is closer to e.g. mechanical hardness or temperature, physical properties that we all take seriously, both in everyday life and

in science. E.g., hardness is considered to be reducible to certain microphysical configurations in a case-by-case fashion, as it is realized differently in different kinds of solid matter. However, as there are no obvious cases of teleology reduction similar to the case of hardness or temperature, the project is more challenging than in the mentioned cases, but it is promising, and that promise could also serve as an argument for taking teleology more seriously.
KEYWORDS: teleology, eliminativism, fictionalism, physicalism, reductionism

GYULA KLIMA
Teleology, Intentionality, Naturalism

This paper argues for the contemporary tenability of a "mentalist, Scholastic-Aristotelian" theory of teleological explanations, *pace* contemporary physicalism/naturalism.
KEYWORDS: Aristotle, intentionality, naturalism, teleology, physicalism

DÁNIEL KODAJ
The Metaphysics of Spooky Teleology

This paper aims to define spooky teleology (the kind of teleology that modern science allegedly exorcised). I argue that understanding this unfashionable concept is important for a variety of reasons, including the improvement of biology education, the reconstruction of certain historical doctrines, and rational debate about teleology in light of modern science. After reviewing a number of possible candidates, I outline and defend a conception based on Richard Braithwaite's notion of plasticity. My proposed definition identifies spooky teleology with irreducible persistent plasticity. I argue that this idea beats all known rivals and it has the potential to foster genuine, empirically informed debate between naturalists and anti-naturalists.
KEYWORDS: dispositions, plasticity, powers, retrocausality, teleology

MOHSEN MOGHRI
An Axiological Ultimate Explanation for Existence

Why is there something concrete rather than nothing? There are many suggestions to explain the existence of our world. But a suggestion can rule out all others that leave no concrete thing unexplained and throw up no further why question. One such ultimate explanation may only be found in something that can carry its own explanation within itself. In this article, I attempt to find one such explanation for all existence. A variant of self-explanation, namely self-subsumption – obtaining of a fact literally in virtue of itself – may succeed in offering an ultimate explanation. One suggestion for a self-subsuming principle is that all possible worlds exist; however, it cannot satisfy us in our deep convictions concerning inductive inferences. Thus, I construct a self-subsuming ultimate explanation that is sufficient to fulfil the latter convictions. My hypothesis is based on the Axiological explanation and suggests that all intrinsically valuable possible worlds are

required to exist; the latter fact also obtains as one of those intrinsically valuable possibilities and therefore is required to exist.
KEYWORDS: axiology, explanation, nothing, teleology, John Leslie, Nicholas Rescher

MICHAEL RUSE
Darwin and Design

Many people, notably Richard Dawkins, author of *The God Delusion*, argue – or are taken to argue – that the chief effect of *The Origin of Species* by Charles Darwin was to finish off Christianity. I shall argue that the story is more complex – and interesting – than this. Darwin's chief achievement was to show how the design-like nature of organisms – the hand, the eye, the heart – can be explained by unbroken law, without direct need of a reference to a Designer, a deity like the Demiurge in Plato's *Timaeus*. Having offered up such an explanation, the way was opened for sound non-belief, although almost always non-believers – agnostics and atheists – take their stance less on science and more on grounds of theology and philosophy.
KEYWORDS: Darwin, design, evolution, teleology

AYUMU TAMURA
The Role of Experience in Descartes' Metaphysics: Analyzing the Difference Between *Intuitus*, *Intelligentia,* and *Experientia*

Descartes uses the term experience (*experientia*; expérience) many time not only in the subject of physics but also in the one of metaphysics, especially in the arguments about the *cogito* and the free will: "he learns ['I am thinking, therefore I am'] from experiencing in his own case that it is impossible that he should think without existing" (*2ae Resp.*, AT-VII, 140; CSM-II, 100); "I cannot complain that the will or freedom of choice which I received from God is not sufficiently extensive or perfect, since I know by experience that it is not restricted in any way" (*Med.*, AT-VII, 56; CSM-II, 39), and so on. However, it is not clear what Descartes means by the term *experientia*; he never defines it. Then what is experience in Descartes' metaphysics? In this paper, I intend to explore what Descartes meant by the term "experience" in the context of metaphysics. To be concrete, I first compare Descartes with earlier philosophers and clarify that Descartes' use of the term "experience" has characteristics that were not recognized earlier (Section 1). I then clarify what the role of experience in Descartes is, while examining the validity of previous studies that equate Descartes' experience with intuition and understanding (sections *2* and 3).
KEYWORDS: Descartes, experience, intuition, understanding