PHILOSOPHICAL Imagination

Thought Experiments and Arguments in Antiquity

EDITED BY

Boris Vezjak

Philosophical Imagination

Philosophical Imagination:

Thought Experiments and Arguments in Antiquity

Edited by

Boris Vezjak

Cambridge Scholars Publishing

Philosophical Imagination:
Thought Experiments and Arguments in Antiquity

Edited by Boris Vezjak

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-6975-6
ISBN (13): 978-1-5275-6975-1

TABLE OF CONTENTS

Part 2: Philosophical Arguments, Imagination and Thought

PREFACE

> Suppose now that there were two such magic rings, and the just put on one of them and the unjust the other. No man can be imagined to be of such an iron nature that he would stand fast in justice. No man would keep his hands off what was not his own when he could safely take what he liked out of the market, or go into houses and lie with any one at his pleasure, or kill or release from prison whom he would, and in all respects be like a God among men. (Plato, *Republic*, translation by Benjamin Jowett)

Historians are not sure if Galileo ever carried out the legendary experiments at the leaning tower of Pisa. Aristotle's theory of gravity stated that objects fall at a speed proportional to their mass. That is, the heavier the object, the faster it falls under gravity. By dropping two balls of different masses, Galileo wanted to demonstrate that their time of descent was independent of their mass. He might not have actually ever conducted such an experiment, yet it became an essential part of the history of physics. The philosopher of science James Robert Brown has called it "the most beautiful thought experiment ever devised"– we normally consider it as a "thought experiment", proving our thesis without actual empirical experimentation.

Thought experiments have been the subject of intense philosophical debate for decades. The term is well known, it is derived from German "Gedankenexperiment" or "Gedankenversuch", first used by the Danish chemist Hans Christian Ørsted in the 18th century. Within contemporary analytic philosophy, they are accepted as a methodological tool used to improve philosophical knowledge and to reconsider or refute philosophical theories. Their fundamental goal is to provide a simplified hypothetical situation that would show the basic assumptions of a particular theory being defended, or try to present evidence that would be in favour or against theories themselves. Over the years, different thought experiments have gained considerable prestige within various philosophical disciplines. They can thus be found within ethics, epistemology, philosophy of science, and are no strangers to other scientific disciplines like physics or mathematics, which only shows their influence and versatility within academia.

In thought experimenting, we perform a special activity of imagining different scenarios in order to test a theory or hypothesis, usually in strong contrast to empirical experiments, where we test theoretical intuitions of sensory observation of objects and events in the external physical world.

Scientific thought experiments appear to be related to actually performed experiments, although the nature of this relationship is not entirely clear. The physicist Ernst Mach envisioned a thought experiment as a necessary prerequisite for every experiment carried out; he said that the experimenter and the inventor must keep an eye on the intended arrangement before actually realizing it. The first obvious distinction between a philosopher and a scientist is that the former is relying on thought experimentation as an original and consistent test method without proving it in the laboratory. Such exclusive preference to thought experimentation is typical for drawing a distinction between the methodology of philosophy and the methodology of natural sciences; the latter can lead to real scientific advances (we need only reflect on the famous thought experiments devised by Galileo, Newton, and Einstein), whereas philosophers address various kinds of ethical and metaphysical issues using intuition or imagination, associating thought experiments with the experimenter's mind. Some authors even believe that philosophy without thought experiments seems unthinkable, and that without thought experiments, even more so than in sciences, philosophy would be profoundly impoverished.

Today, two thought experiments developed by Edmund Gettier (1963) are considered as a paradigmatic example of traditional epistemology. According to the traditional view, our knowledge is a justified true belief. Well, Gettier claimed that his thought experiments show something else: if there is a philosophical tradition that understands knowledge of a proposition as a justified true belief in that proposition, Gettier proved, convincingly, that in some situations we have a belief that is both true and well supported by evidence, yet utterly fails to count as knowledge. One of the greatest epistemological discoveries on the very nature of knowing, written in a three-page article by an unknown philosopher, started various research programs and continues to influence debates in epistemology to this very day. Other paradigm thought experiments in philosophy include famous examples made by other philosophers, such as Philippa Foot's (1967) and Judith Jarvis Thomson's (1976) trolley experiments, Hillary Putnam's twin earth experiment (1975), John Searle's Chinese room experiment (1980), Saul Kripke's Schmidt/Gödel experiment (1980), and Frank Jackson's neuroscientist experiment (1982), all of which have had enormous influence in different philosophical schools.

But how far back can we go to track thought experiments in history of philosophy and science, and who was the first to conceive and discover them? They seem to be used both in philosophy and science in the distant past of antiquity. Archytas of Tarentum, for example, tried to prove the infinity of the universe with his probably first thought experiment in the

Western tradition. His cosmological view is still regarded as one of the most compelling arguments ever produced in favour of the infinity of space; later, it was taken over and adapted by the Stoics and Epicureans, especially Lucretius, who claimed that if the universe is finite, then it must be surrounded by a final boundary. But no boundary can be final, because there always has to be something on the other side; hence the universe must be infinite.

Thought experimenting by ancient philosophers is often open to debate: in what sense did their reasoning really concern thought experimentation? Sometimes the authors apparently support philosophical theories, in other cases experiments are carried out to reject some philosophical ideas; and yet in other cases the philosophers can only propose suspension of judgement. In Plato's *Republic*, Glaucon uses the myth of Gyges to demonstrate why people who practice justice do so unwillingly. A challenge, posed to Socrates and provided through some sort of thought experiment by imagining the effects of using the ring of invisibility, was intended to answer the question about human nature and our basis for the inclination towards justice or injustice. The example of the third kind of thought experiment was developed by Sextus Empiricus in his work *Against the Physicists*, where he attacked the famous Epicurean doctrine of the existence of atoms while discussing the possibility of motion, proving that, if motion is possible, there would be no partless bodies, and hence atoms do not exist.

The present book seeks to add new insights to the otherwise not too extensive literature on the beginnings of thought experiments in antiquity. The idea to publish this book arose during the international conference organized by the Department of Philosophy of the Faculty of Arts in Maribor ("Philosophical Imagination, Thought Experiments and Arguments in Antiquity", Maribor, 9-10 October 2018). The reader will notice that the book is divided into two chapters, because it is sometimes extremely difficult to distinguish between a general discussion on hypothetical reasoning or imagination and thought experiments in a very strict sense. The present reading of selected authors therefore seeks to deepen the current, otherwise scarce discussions of whether it is possible to articulate a discussion about thought experiments and about its arguments from the historical perspective of philosophy and science. It may sometimes seem that, in a loose sense, any philosophical reflection can already be interpreted as some form of thought experimentation. Although its functions are very diverse and complex, and often closely linked to other cognitive tools, such as visualization, imagination, or idealization, the contributions collected in this book may provide new insights into how the concept of a thought experiment coincides with more modern perspectives.

P.S. Finally, I would like to thank my dear friends and colleagues from the Department of Philosophy in Maribor for all the unconditional support, and a special thanks to my colleague Tadej Todorović for his help and support in preparing this manuscript.

Boris Vezjak, editor

List of Contributors

Andre M. Archie specializes in the History of Ancient Greek Philosophy and Ancient Greek Political Philosophy. His latest research focuses on methodological issues, i.e. styles of argumentation in Plato and Aristotle. He has published in various journals such as the *Journal of Philosophical Research*, *Ancient Philosophy*, *History of Political Thought*, and *Scholia*. Andre is the author of *Politics in Socrates' Alcibiades: A Philosophical Account of Plato's Dialogue Alcibiades Major* (Springer, 2015). He is currently working on a book manuscript that is under contract with The Catholic University of America Press titled: *Socratic Conservatism, Socratic Questions: The Right Turn in Plato's Political Dialogues*.

Predrag Cicovacki is Professor of Philosophy at the College of the Holy Cross (USA), where he has been teaching since 1991. He served as a visiting professor in Germany, France, India, Luxembourg, Russia, and Serbia. He authored ten books, the latest of which are: *The Luminosity of Love* (2018) and *Gandhi's Footprints* (2015). He edited eight books, which include: *Tolstoy and Spirituality* (2018; with Heidi Grek) and *Nonviolence as a Way of Life* (2017; with Kendy Hess).

Matteo Cosci is research fellow at Ca' Foscari University of Venice, Italy. He is author of *Verità e comparazione in Aristotele* (Venice, 2014) and co-editor (with M. Sgarbi) of *The Aftermath of Syllogism* (London - New York, 2018). He is the Assistant Editor of the online journal *Philosophical Readings* and a recipient of the "Marie Skłodowska Curie-IF". His research topics include Aristotle, the history of Aristotelianism, and Galileo Galilei.

Milan M. Ćirković is a research professor at the Astronomical Observatory of Belgrade, (Serbia) and a research associate of the Future of Humanity Institute at the Oxford University (Oxford, UK). His primary research interests are in the fields of astrobiology (habitable zones, habitability of galaxies, SETI studies), philosophy of science (futures studies, philosophy of physics, philosophy of cosmology), and risk analysis (global catastrophes, observation selection effects, epistemology of risk). He co-edited the widely-cited anthology on *Global Catastrophic Risks* (Oxford University Press, 2008, with Nick Bostrom), wrote three monographs (the latest being

The Great Silence: The Science and Philosophy of Fermi's Paradox, Oxford University Press, 2018), as well as four popular science/general non-fiction books, and authored about 200 research and professional papers.

Drago Đurić is Professor of Philosophy at the Department of Philosophy, University of Belgrade (Serbia). He has published four books and papers in a number of philosophical journals. The latest book is titled *Eternity of the World in Antiquity and the Early Middle Ages* (written in Serbian–Belgrade, 2015). His work is focused on the research of ancient philosophy and science, early medieval thought, philosophy of religion, metaphysics, especially the metaphysics of free will. Current research concerns the preparation of a book under the working title: *Ionian Physics: Change and Matter.*

Maja Malec is Assistant Professor of Philosophy at the Faculty of Arts, University of Ljubljana. Her main research interests include metaphysics, epistemology, and philosophy of logic.

David Merry is an independent scholar specialising in connections between ancient dialectic and ancient ethics. He holds a PhD from the Humboldt University in Berlin, and has held research positions at the Humboldt University Berlin and Heidelberg Medical Hospital.

Nenad Miščević is Professor of Philosophy at the Faculty of Arts, University of Maribor. He has been working on thought experiments in various areas (published in journals like *International Studies in the Philosophy of Science, Acta Analytica* and *Croatian Journal of Philosophy*). His special interest is in political thought experiments, about which he has written the article "Political Thought Experiments from Plato to Rawls" in *Thought Experiments in Science, Philosophy, and the Arts* (edited by Frappier and Brown) and a chapter on political thought experiments in *The Routledge Companion to Thought Experiments*. He is currently preparing a book on political thought experiments.

Marek Picha focuses on epistemology and philosophical methodology. Lately he has been researching argumentative moves in debate about epistemological problem of generality, and a persuasive usage of appeal to ignorance in Descartes' *Meditationes*. Marek is the author of *What if* (2011) and *100 Thought Experiments in Philosophy* (2013) [both in Czech].

Danilo Šuster is Professor of Philosophy, teaching logic, epistemology and metaphysics at the Department of Philosophy, Faculty of Arts, University of Maribor. He is Editor in Chief of *Acta Analytica* (Springer) and author of several books on logic and analytic philosophy in general (written in Slovene language). His publications include articles on free will and causation.

Matjaž Vesel specializes in the field of the history of medieval, renaissance and early modern science and philosophy. His latest research focuses on some fundamental conceptual, methodological, epistemological and theological aspects of the Scientific Revolution as conceptualized in the works of Nicholas of Cues, Copernicus, Kepler, Galileo, Descartes, Newton and others. He has published numerous articles and monographs. His latest monograph, titled *Copernicus: Platonist Astronomer-Philosopher: Cosmic Order, the Movement of the Earth, and the Scientific Revolution* (Peter Lang, 2014), has received special attention. He is currently working on a book manuscript concerning Newton's criticism of Descartes' metaphysics and natural philosophy.

Boris Vezjak is Associate Professor of philosophy at the Department of Philosophy, Faculty of Arts, University of Maribor, Slovenia. His professional fields of interest cover various subjects, such as history of philosophy, discourse theory, media analysis and theory of argumentation. His publications include many articles, translations (*Plato's Parmenides*, *Plato's Charmides*, *Plato's Philebus*) and seven monographs, mainly focused on the history of philosophy, media and social analysis.

PART 1:

THOUGHT EXPERIMENTS FROM PLATO TO HUME

CHAPTER ONE

THE ANATOMY OF THREE THOUGHT EXPERIMENTS IN PLATO'S REPUBLIC, APOLOGY, AND IN ALCIBIADES MINOR[1]

ANDRE M. ARCHIE

Introduction

The standard view of thought experiments, and their role in the advancement of science and philosophy, is expressed in Thomas S. Kuhn's essay, "A Function for Thought Experiments" (Kuhn 1977). However, Kuhn's explanation of thought experiments underemphasizes how they simultaneously reform an agent's concept and actions; how physical verisimilitude manifests itself.

I remedy Kuhn's underemphasis by citing three prominent examples of thought experiments from the *Republic*, *Apology* and *Alcibiades Minor*. My main purpose in juxtaposing Kuhn's and Plato's illustrations of the function of thought experiments is to show that Plato seems to anticipate and go beyond Kuhn's concern that thought experiments satisfy the condition of physical verisimilitude. Plato's thought experiments demonstrate that thought experiments are not conducted merely to alter the conceptual apparatus of the interlocutor on the explored topic, but rather to alter the interlocutor's conceptual apparatus for the sake of altering his actions (i.e. practical rationality). Thus, Plato's concern that theory not be separated from practice is a modern concern. Although there are other ancient texts that contain thought experiments, Plato's discussion of thought experiments reminds us of the endurance and relevance of Plato's methodology.[2]

---

[1] This is a revised version of the paper "The Anatomy of Three Thought Experiments in Plato's Republic, Apology, and Alcibiades Minor", published in *Journal of Philosophical Research*, Vol. 35: 305-321, 2010.

[2] Ierodiakonou provides several examples of ancient thought experiments, but her focus is on ancient thought experiments as a tool for refuting rival philosophical

The outline of my argument is as follows. In section I, I discuss Kuhn's formulation of the function of thought experiments. In section II, I preview the anatomical features of each thought experiment in all three dialogues, and how these features fulfil and go beyond Kuhn's requirements. In section III, Glaucon's Ring of Gyges thought experiment is discussed along with Socrates' response to it. Socrates builds the city for pigs as an alternative thought experiment. His eudaimonistic argument for the city of pigs offers incentives for Glaucon to live a happier life. Thus, refuting Glaucon's belief that moral behaviour is contrary to one's self-interest. In section IV, Socrates' self-reflective thought experiment during his trial is shown to be a useful way to get the Athenians to appreciate the nature of his unique ambition. Socrates' refusal to cease practicing philosophy illustrates the gap between Socrates' priorities and the Athenians'. However, Socrates' refusal also illustrates practical ways in which the Athenians can order their disordered priorities. In section V, Socrates helps a young, ambitious Alcibiades become aware of his ignorance by posing a thought experiment. The lesson that Socrates conveys to Alcibiades through the thought experiment is that there is no way to distinguish between precious and pernicious things without knowledge. Such a distinction is made possible by recognizing that practical knowledge serves as the basis of virtue.

I. Thomas Kuhn and Thought Experiments

In "A Function for Thought Experiments", Kuhn surveys what he calls the "mainstream" view of the function of thought experiments, and discusses several arguments to correct what he sees as its main deficiencies. The mainstream view, characterized by Kuhn, holds that understanding produced by thought experiments is not a better understanding of nature, but rather a better understanding of the scientist's conceptual apparatus:

> On this analysis, the function of the thought experiment is to assist in the elimination of prior confusion by forcing the scientist to recognize contradictions that had been inherent in his way of thinking from the start. Unlike the discovery of new knowledge, the elimination of existing confusion does not seem to demand additional empirical data. Nor need the imagined situation be one that actually exists in nature. On the contrary, the thought experiment whose sole aim is to eliminate confusion is subject to only one condition of verisimilitude. The imagined situation must be one to which the scientist can apply his concepts in the way he normally employed them before. (Kuhn 1977, 242)

doctrines. She says very little about the relationship between thought experiments and action. See Ierodiakonou 1991 and Rescher 1991.

Kuhn takes issue with the mainstream view, which ignores thought experiments that operate according to a nonstandard inconsistency. Such inconsistencies, according to Kuhn, result from a peculiar loyalty to the analytic/synthetic distinction (Kuhn 1977, 255-259).

He argues that thought experiments make explicit a concept's inconsistencies. However, these are not inconsistencies typically identified by logicians. Kuhn contrasts the standard concept of inconsistency, e.g. square circle, with a non-standard example, e.g. faster. To illustrate non-standard inconsistency, he cites Piaget's laboratory situation in which children are presented with moving cars for the purpose of making explicit inconsistencies in their goal-reaching criteria (see Piaget 1946, Ch. 6 and 7). Kuhn posits that non-standard inconsistencies are ignored because scientists believe that thought experiments yield only analytic propositions. Consequently, the mainstream view holds that thought experiments are not informative; they provide no knowledge about reality.

Historically and philosophically, the argument Kuhn offers to overcome the analytic/synthetic distinction in the assessment of the function of thought experiments is significant, but beyond the scope of what I find most illuminating in Kuhn's essay.[3] My focus will be directed at Kuhn's reformulation of the mainstream view that the only condition of verisimilitude which thought experiments should be subjected to is the condition that concepts in the thought experiment be applied the same way they were prior to such an experiment.

Kuhn's basic argument is that, in addition to satisfying the condition of logical verisimilitude, i.e. internal consistency, thought experiments must satisfy the condition of physical verisimilitude. Thought experiments must teach the scientist or philosopher about his concepts and the world together. Effective thought experiments reveal the discrepancy between the phenomena and the scientist's or philosopher's understanding of the phenomena. In other words, nature and conceptual apparatus are jointly implicated (Kuhn 1977, 265).[4] In saying that an agent learns about the world through the joint implication of concepts and nature, we presume Kuhn is referring to the

[3] For an informative discussion of Kuhn and thought experiments, see Sorensen 1992, 112-131.

[4] E.g. Kuhn cites Galileo's *Dialogue concerning the two Chief World Systems* to illustrate that, for Galileo's readers, in learning about the concept of speed, they simultaneously learn how bodies move (Kuhn 1977, 253). Galileo's thought experiment illustrates that uniform horizontal motion does not affect the outcome of localized experiments. Outside of the localized context, effects can be quantified. On Galileo's thought experiment, and the role of thought experiments in science and philosophy, see Cohen 2005.

precision and scope gained in the agent's actions in the world brought about by robust concepts.[5]

I am in agreement with Kuhn that most people learn about their concepts and the world together, but the effects of thought experiments on an agent's actions are given very little attention in his reformulation of the mainstream view. The principle entailed by my position, and which justifies my interpretation of Plato's thought experiments, is that virtue is a property of the intellect. What alone can make an interlocutor go wrong is ignorance. Plato's thought experiments are crucial in getting the interlocutor to straighten out his views, thus helping him achieve the good for which he aims.

Given my intellectualist moral account of the effects of Plato's thought experiments on action, linking Kuhn's discussion of thought experiments in the natural sciences to Plato's may appear to be odd. However, the commonality between thought experiments in the natural sciences to those in moral philosophy is that they both allow the experimenter to grasp the law-like structures of nature or of morality. Such structures operate like Plato's Forms, and the right sort of thought experiments facilitate the perception of abstract intuitions. Although in each of Plato's thought experiments the experiment's empirical premises take us beyond sense experience, Plato makes it clear that the intuitions his thought experiments give rise to are inseparable from practical rationality.[6]

II. Anatomical Features of Plato's Thought Experiments

The three Platonic thought experiments I have chosen to focus on contain three features. These anatomical features are efficiency, conceptualization, and refutation. They fulfil and go beyond Kuhn's requirements for thought experiments by emphasizing practical rationality. The following description of the features should help in framing how to construe Plato's thought experiments and to show that their effectiveness is primarily seen in the reformation of behaviour.

a. Efficiency: the supposition that each thought experiment sets out from is in principle unrealizable. Consequently, there is no recourse

[5] The accuracy of this presumption is confirmed by Kuhn's admission that "the effects [of thought experiments] are much closer to those of actual experimentation than has usually been suggested" (Kuhn 1977, 242).

[6] For some interesting ideas on the relationship between thought experiments in the natural sciences and in moral philosophy, see Brown 2004.

to technical apparatuses or laborious procedures. What is needed is pure supposition bolstered by vivid imagery.[7]

b. Conceptualization: each thought experiment assumes that the interlocutor's problem is not at the level of observing the phenomena (e.g. human nature, priorities, and sensible particulars); but rather at the level of conceptualizing the phenomena. In the dialogues, understanding the phenomena appropriately has practical consequences.
c. Refutation: each thought experiment functions to refute a rival supposition. Such a refutation takes the form of an implied conditional argument:

(Antecedent)	A→B	(Consequent)
	A	
(Conclusion)	Therefore, B.	

This sort of argument shows that a rival's supposition could be falsified.[8] Consequently, continued belief in the supposition by the rival may lead to negative consequences.[9]

---

[7] Some commentators have argued that thought experiments are just arguments that make use of colourful or memorable illustrations or situations. For example, this is what John D. Norton argues (in his 2004 essay, "Why thought Experiments do not Transcend Empiricism") Norton's thesis is stated as follows:

> (1a) (Context of justification) all thought experiments can be reconstructed as arguments based on tacit or explicit assumptions. Belief in the outcome-conclusion of the thought experiment is justified only insofar as the reconstructed argument can justify the conclusion.
>
> (1b) (Context of discovery) The actual conduct of a thought experiment consists of the execution of an argument, although this may not be obvious, since the argument may appear only in abbreviated form and with suppressed premises. (Norton 2004, 50)

For Norton, thought experiments are merely argument and inference (i.e. thought experiments are just philosophical arguments that carry us from perception to proposition). From the same collection of essays, the alternative claim that thought experiments are not just colourful philosophical arguments is advanced by Brown, in his 2004 essay "Why Thought Experiments Transcend Empiricism".

[8] "Rivals" should be interpreted broadly to include interlocutors and societal endoxa (conventional opinions).

[9] These anatomical features are not limited to Plato's thought experiments, although they illustrate the mechanics of thought experiments in the dialogues. It is important to note that the scholarly literature on thought experiments refers to some version of the features I have described as efficiency, conceptualization, and refutation. See Burge 2007; Glymour and Wimberly 2007; Gendler 2000; Rescher 2005; Swirski 2007.

III. The Ring of Gyges (360b2-369c)

The thought experiment presented by Glaucon in Plato's *Republic* is the second of two arguments designed to explore the nature of morality. Glaucon's experiment is supposed to lend credence to the claim that just actions are solely motivated by fear of punishment.[10] To support his claim, Glaucon recounts the tale of the Ring of Gyges, which is about a ring with the power to make its wearer invisible:

> Let's suppose that there were two such rings, one worn by a just and the other by an unjust person. Now, no one, it seems, would be so incorruptible that he would stay on the path of justice or stay away from other people's property, when he could take whatever he wanted from the marketplace with impunity, go into people's houses and have sex with anyone he wished, kill or release from prison anyone he wished, and do all the other things that would make him like a god among humans. Rather his actions would be in no way different from those of an unjust person, and both would follow the same path. This, some would say, is a great proof that one is never just willingly but only when compelled to be. (*Republic* 360b2-360c4)[11]

The power of Glaucon's beliefs, as expressed by his thought experiment, derives from its lack of concern with actualized beliefs (beliefs presently at work in our thinking of them) but with dispositional beliefs (beliefs that would dictate actions in the appropriate circumstances). Glaucon's thought experiment is grounded on the assumption that human desires are naturally unlimited and that ethics is rightly non-eudaimonistic.[12] Eudaimonism is a

[10] Glaucon defends injustice in order to show Socrates the sort of defence he wants of justice. The vividness of Glaucon's defence elicits Socrates' ironic concern (361d3-5; see 368a4-c) that Glaucon may have been too vigorous a defender of injustice.

[11] Except where indicated, all translations from the *Republic* are from G.M.A Grube's translation (Plato 1997, edited by Cooper).

[12] The limited / unlimited distinction is a crucial feature of the Socratic Method. Such a distinction is based upon the belief that in order to gain a comprehensive understanding of things all possibilities must be considered. Through thought experiments, Plato goes about considering the possibilities by removing all unity or limitations from things. What is left is pure unlimitedness. Individuals like Glaucon slide into relativism and skepticism by supposing that there is just unlimitedness (see Plato's *Theaetetus* for the depiction of this view). In dealing with souls, there is danger in unlimited, ceaseless desires. However, virtue provides a limit principle and is a necessary part of a well-shaped character. A happy life requires putting some good limits on desires or appropriately shaping our characters. A just community requires giving the polity a good configuration or constitution.

paradigm for reflecting on the relationship between virtue and happiness.[13] Its style of reflection is not so much about a criterion for choosing the best action, but about the best way to live. It is person centred rather than act centred. Although eudaimonism need not make metaphysical assumptions, healthy eudaimonism depends upon the belief that the universe generally supports limited human desires as a necessary condition for the possibility of human happiness.

Glaucon's non-eudaimonistic position posits that: (1) humans are always comparing their situations with others, i.e. they inevitably try to outdo one another, and this makes human desires ceaseless and accelerating. Consequently, no quantity of goods like honour, money, and power suffices if humans are always comparing themselves. And (2) there is nothing fixed that humans always want and nothing of any kind can fully satisfy them, so there is no determinateness in what people desire. No kind or quality of things is sufficient. The name for the condition of outdoing others for the sake of gain is called pleonexia. It is opposed to the condition of equality, which may be either numerical or proportional equality under the law. Contrary to what the conventional belief of justice actually holds, Glaucon asserts that we will see pleonexia if we give to both just and unjust the ring of invisibility. The ring will disclose natural desires and show that hardly anyone is voluntarily just. Glaucon's argument could be analysed in the following way:

> Premises:
> (1) The desire for undue gain (i.e., pleonexia) is a desire of most people. Such desires manifest in those unrestrained by a lack of power.
> The subject matter of the first premise can be referred to as the Phenomena (*what* is to be explained).
> (2) Any desire most people manifest when unrestrained by lack of power is a desire that is natural for humans.
> The subject matter of the second premise can be referred to as the Principle of Interpretation (how we *ought* to explain human nature given the Phenomena).
> Conclusion:
> The desire for undue gain is a desire that is natural for humans.

Conceptual Reform: Socrates, Glaucon and Eudaimonism

Socrates refutes both of Glaucon's premises by reformulating his concept of the Phenomena (and by implication Glaucon's Principle of

[13] On the types of eudaimonism see Vlastos 1991, 200-232.

Interpretation).[14] The premise of the refutation is based on what humans are like originally; what humans are like in the city for pigs.[15] Socrates argues that only under faulty circumstances will people desire undue gain, and that the best people must be the standard. His refutation of Glaucon's argument makes explicit the joint implication between the concept of the Phenomena and nature. If Socrates is correct, he will have shown that human desires are limited and that human desires are in principle satisfiable. Consequently, the virtuous life is possible and sustainable for the individual and society as the best type of life.

Socrates' thought experiment begins by establishing that cities start because humans lack self-sufficiency but need many things.

> Therefore our citizens must not only produce enough for themselves at home but also goods of the right quality and quantity to satisfy the requirement of others. –They must. –So we'll need more farmers and other craftsmen in our city. –Yes. And others to take care of imports and exports. –And they're called merchants, aren't they? –Yes. –So we'll need merchants, too. (*Republic* 371a-11)

As a result of each citizen fulfilling their social roles, desires can be satiated. Socrates explains:

> First, then, let's see what sort of life our citizens will lead when they've been provided for in the way we have been describing. They'll produce bread, wine, clothes, and shoes, won't they? They'll build houses, work naked and barefoot in the summer, and wear adequate clothing and shoes in the winter. For food, they'll knead and cook the flour and meal they've made from wheat and barley. /…/ They'll enjoy sex with one another but bear no more children than their resources allow, lest they fall into either poverty or war. (*Republic* 372a3-c)

---

[14] My discussion of each thought experiment takes it for granted that what is being reformed is how the interlocutor sees the topic under consideration in the respective dialogues. Typically, the interlocutors' seeing is assumption-laden, and it is what accounts for the hubris and hard-headedness that animates their characters. Consequently, Socrates is not merely helping interlocutors see better; he is helping them see the best way to live. On the relationship between seeing as a physical state and seeing as an experience, see N.R. Hanson's classic text *Patterns of Discovery* (1958, 4-30).

[15] The building of the cities (city for pigs, the luxurious city, reformed city, and the philosophical city) in the speech is one big thought experiment countering Glaucon's Ring of Gyges thought experiment. However, it is within the city for pigs that Socrates establishes the principle on which other cities either deviate or conform: our most basic desires in life are limited and determinate.

By insisting that self-sufficiency and satisfaction are possible in the city for pigs, Socrates is suggesting that human desires are limited in principle. Consequently, there is no conflict due to competition in this city, because the citizens just naturally do what their natural abilities prescribe. Even within the intimate realm of reproduction there is no conflict between public and private good. Although women are not mentioned, sexual desire is separated from the desires that lead to too many children.

Socrates' thought experiment is a frontal attack on Glaucon's assumption that human desires are unlimited. It is significant, too, that, for Socrates, the city begins because of human neediness rather than fear as Glaucon's thought experiment suggests.[16] Glaucon sees the political order as a conventional arrangement; a collective agreement entered into by weak people to compensate for their individual weaknesses.

Socrates' position regarding human desires is eudaimonistic. Unlike Glaucon, Socrates does not view human desires as a threat, since people are to be supplied with necessities, and they need not be in competition with others for them. Consequently, the virtuous life secures the orientation towards the right thing to do but also the doing of a righteous act through the proper motivation. Socrates' eudaimonism posits that:[17] (1) Humans may compare situations, but they can be content and keep to their own. Love of their own is compatible with the common good. They do not have to outdo others but can mind their own business. (2) Desires seem indeterminate when they are all mixed up, such as "money love" becoming confused with "honour love", so that they drive each other on. (The miser may want more money but primarily seeks to preserve what he has, or the pleonexic person gets honour from accumulating ever more wealth.) However, when desires are separated, they are all satisfiable in principle.[18]

For Socrates, natural human needs lead to community, so the community is natural for humans inasmuch as only in it can natural desires receive satisfaction. The city for pigs that Socrates builds is intended to satisfy the most necessary human needs. Such needs are food, shelter, and clothing. Desires are (only) for what we need. Thus, desires are determinate. Socrates is also separating basic, human desires from higher, more complex desires.

---

[16] Premise (1) and (2) of Glaucon's argument is the summation of his earlier (358e-359b) claim that most people follow the law not as something inherently good but due to their weakness, though they still desire injustice naturally. Conversely, those who are strong enough should then do injustice when they can get away with it.

[17] Socrates' eudaimonistic claims are a direct refutation of the claims Glaucon endorses (see above).

[18] Justice seems to be each doing his own task and the natural need of each for the other leads to the founding of the city. See 372a-c.

Socrates' concept of the Phenomena (desires) posits no original disorder; rather, human needs enjoin the natural order of the human community. Socrates' argument can be analysed in the following way:

> Premises:
> (1) Just, "necessary" desires, i.e. limited, ordered, satiable, beneficial desires, are the desires that the truly best persons have manifest; most other persons can be educated in a good society to have appropriate desires (the Phenomena).
> The first premise claims that desires are found and perfected within the community. The human capacity for logos and action governed by logos can be fully realized only within the community.
> (2) The desires of the truly best persons are natural for human desires (Principle of Interpretation).
> The second premise claims that individuals within a community governed by laws and justice are the best type of humans. Socrates reads human nature off from how individuals are when most fully developed rather than how individuals are initially.
> Conclusion:
> Just, "necessary" desires are natural for humans.

Given what we have said about Glaucon's thought experiment and Socrates' response to it, Plato seems to suggest that how one conceives the nature of human desires (the Phenomena) determines how one will live. This view of human psychology is eudaimonistic and it holds that what alone can cause error—moral and intellectual—is ignorance. On this view, Platonic thought experiments always seem to satisfy the condition of physical, i.e. practical, verisimilitude: one must always strive to live the virtuous life. Whereas the relationship between conceptual reform and its effect on practical activity is underemphasized by Kuhn, Plato's thought experiments are often refutations of an interlocutor's faulty conceptualization of the topic under discussion. Refutations are designed to straighten out views, engender self-knowledge, and help the interlocutor achieve the good for which he ought to aim. Glaucon's argument for the nature of desires is subjected to refutation. His descriptive account of desires naturalizes how humans are when in competition and conflict with one another to satisfy their pleonexia. In so far as justice is concerned, Glaucon cannot move beyond the conventional origins of justice as a way of limiting conflict within the community. Socrates' thought experiment seeks to reform Glaucon's concept of desires. Such reformation seeks not merely to alter Glaucon's conceptual apparatus; rather it offers the incentives to live a happier life. Once the nature of happiness is determined, we should seek what contributes to happiness. Socrates' prescriptive account of desires

seems to assume that unless Glaucon allows for an end or limit to the satisfaction of desires, it will be hard to find true happiness.

Socrates may be optimistic, but his goal is nonetheless to get Glaucon to see that he is rational and rationality is expressed through him aiming at what is good. Thus, refuting Glaucon's belief that moral behaviour is contrary to one's self-interest. The compelling aspect of Socrates' thought experiment is that it is not concerned so much with Glaucon's actual beliefs as it is with equipping Glaucon with the right dispositional beliefs (beliefs that would dictate the right actions in appropriate circumstances). If Socrates can win Glaucon over to considering his eudaimonistic position, he has at least gotten Glaucon a step closer to reforming his behaviour.

IV. Socrates' Self-Reflective Questioning (29c6-e3)

The thought experiment presented by Plato in the *Apology* is designed to show the Athenian jurors the degree to which Socrates is willing to face death in promoting the philosophical life. The contrast between what the jurors offer Socrates as a condition for acquittal and his response seems designed to induce the jurors to alter their behaviour by reconceptualising their understanding of the soul's relationship to the body. The importance of reconceptualising the relationship facilitates Socrates' ultimate goal of having the jurors arrange their priorities in the correct order. The context is Socrates' refusal to stop doing and saying what he supposes is the best thing for him to do. Fear of death will not keep him from philosophizing and cross-examining himself and other people. This leads up to Socrates self-reflectively entertaining the counterfactual choice of being acquitted by the Athenians on the condition that he ceases investigating and practicing philosophy or be put to death:

> Socrates, we do not believe Anytus now; we acquit you, but only on condition that you spend no more time on this investigation and do not practice philosophy, and if you are caught doing so you will die;' if, as I say, you were to acquit me on those terms, I would say to you: "Gentlemen of the jury, I am grateful and I am your friend, but I will obey the god rather than you, and as long as I draw breath and am able, I shall not cease to practice philosophy, to exhort you and in my usual way to point out to any one of you whom I happen to meet: Good Sir, you are an Athenian, a citizen of the greatest city with the greatest reputation for wisdom and power; are you not ashamed of your eagerness to possess as much wealth, reputation

> and honors as possible, while you do not care for nor give thought to wisdom or truth, or the best possible state of your soul?" (*Apology* 29c6-e3)[19]

Socrates' thought experiment presented as a hypothetical choice is purely suppositional. It also illustrates the type of categories and choices Socrates is concerned with in other dialogues.[20] The category under which Socrates' counterfactual choice is offered is The Call of Ambition.[21] Under this category, the hypothetical choice entertained by Socrates is concerned with establishing as fact, for the jurors' consideration, that he is ambitious, but ambitious for the sake of promoting moral reform. When I speak of ambition, I mean those overriding emotions or desires causing an individual to act in a particular way repeatedly to achieve a particular end. I am not speaking of petty desires that are strongly felt and, as a result, cause Socrates to act. Socrates' ambition is *grand* and *noble*.[22] Similar to other dialogues where Socrates confronts interlocutors with hypothetical choices to assess how strongly or to what degree they hold a particular belief, the choice he offers to himself shows the jurors how strongly he believes in his philosophical mission, and in the redemptive nature of philosophy.

Understanding Socrates' Ambition

The distance Socrates is willing to go to reform the jurors understanding of the soul/body relationship is the measure of his ambition and the confidence he has in his philosophical mission. The lack of these greater things, which would demand that he remain quiet and content with the status quo, would be a peculiar kind of death by proxy for Socrates. A belief entails choices, and the choice Socrates self-reflectively offers himself reveals to the jurors his preference for hierarchically ordered worlds according to their perceived value in facilitating his ambition to promote moral reform.

---

[19] Except where indicated, all translations from the *Apology* are Grube's translations in *Five Dialogues* (2002)

[20] It has been argued that there are three distinctive categories under which choices are offered in the dialogues: The Call of Ambition, The Limits of Ambition, and The Transparency of Ambition. E.g. see Maurice 2006.

[21] Thought experiments, like counterfactuals, have been discussed in relation to the variety of Socratic refutations but not as a technique to measure the extent to which an interlocutor is ambitious. See, for example, Carpenter and Polansky's "Variety of Socratic Elenchi" (2002). In the *Apology*, Socrates often assumes the role of the interlocutor. For a denial that the Apology is a dialogue, see Burnyeat's "The Impiety of Socrates" (1997).

[22] After all, the Delphic oracle seems to sanction his ambition. See 20d-e.

Schematically, the counterfactual question Socrates imagines the jurors asking him might look like the following:

> Would you rather have world α (acquittal) in any event; or world β (to die) if ρ ("on condition you spend no more time on this investigation and do not practice philosophy") is true and world γ (implying Socrates is allowed to investigate and practice philosophy) if ρ is false?[23]

If Socrates were certain that ρ was true, he would then choose, as if no conditions were attached, between α and β. Socrates suggests that he would most likely choose β, and he does (see 29e1-42a3). Despite the fact that Socrates will ultimately be put death for choosing as he does, the degree of Socrates' commitment to the life of philosophy helps others initiate the process of *at least* thinking about the question of what constitutes a moral life. Such a process continued even after Socrates' death.[24]

Conceptual Reform: Priorities

The conceptual reform that Socrates' thought experiment seeks to initiate is one that helps the Athenians to be ever vigilant in ordering all that they care about and to have the appropriate priorities in mind. In reform of this type, the jurors' understanding of their priorities and how they live their lives are jointly implicated. What might this ordering of priorities look like? Consideration of a related passage at 29e-30b reveals Socrates' standpoint in challenging the citizens of Athens for neglecting the right order through placing greater value on their personal possessions than on their souls, and thus attaching little importance to the most important things (e.g. wisdom, truth and the soul), while cherishing inferior things (e.g. wealth, reputation and the body). Socrates endorses the following claim:

> Wealth (*chremata*) does not bring about excellence, but excellence makes wealth and everything else good for men, both individually and collectively. (30a9-10)

---

[23] In formulating Socrates' counterfactual question this way, I have benefited from Ramsey's article "Truth and Probability" (1978). Ramsey's article sets out a method to measure degrees of beliefs and other psychological variables through their causal property, which is the extent to which individuals are willing to act on what they believe given hypothetical circumstances.

[24] See Plato's Seventh Letter for a depiction of Socrates' influence on Plato's youthful ambition.

We can interpret this passage in several ways.[25] Either virtue makes wealth and other things good for humans collectively or privately; or virtue does not come from wealth, but from virtue comes wealth and all other goods for man collectively and privately. The ambiguity may be intentional to allow for both interpretations, but each interpretation is anchored in an order that prioritizes three types of goods: goods of the soul; goods of the body; and external goods. Goods of the soul revolve around the mutually entailing ideas of knowledge and virtue; goods of the body include qualities such as health and strength; and external goods include wealth and honours. Socrates believes in the greater value of the soul than of the body and its possessions. What Socrates suggests is that only the goods of the soul allow one to use the other goods well. Even if external goods are most necessary, they are not the highest since the soul is what uses the others. For example, when we consider that it may be true that the virtuous person is rich, presumably it is because such a person knows how to make do or he does the best with what he has due to the moderating influences of the soul.

Given Socrates' sentiments in the passage under consideration, it is clear why Socrates admonishes the Athenians as single-mindedly as his thought experiment attests. The very things that give Athens the reputation for "both wisdom and power" blind it and make it "sluggish" (30e4) with respect to the most important things (30d4). Athens is blinded by its bodily goods and its possessions. Consequently, the right ordering of Athenian priorities becomes the concern of Socrates' investigation and refutation of his fellow citizens.

Following the jury's guilty verdict, Socrates again addresses the issue of Athenian priorities, and what role he played as a private citizen in trying to convince others to concern themselves with the state of their soul as opposed to the body and its possessions. Socrates explains that his counter-assessment must be commensurate with a life that has not been lived quietly or concerned with what occupies the majority of Athenians: wealth, household affairs, and political offices (36b-c). The life that Socrates has tried to live is a life that has been useful, both to himself and to others:

[25] On the various interpretations of the passage in light of the role played by the word *chremata*, see de Strycker and Slings 1994, 138-141. Burnyeat construes the passage as contributing to a larger discussion of the approach to moral philosophy exhibited by the Socratic concern for vice, virtue, and character or (being), and the modern concern for methodology and actions or (doing). For Burnyeat, *chremata* is not money simply, it is valuable possessions in the broadest sense of the word. Thus, virtue (being) is coupled with actions (doing) due to its capability of "dominating and organizing the whole pattern of a man's life" (Burnyeat 1980, 210). Burnyeat's reflections reinforce the main lines of thought we have found in the passage. See "Virtue in Action" (Burnyeat 1980).

> I did not follow that path that would have made me of no use either to you or to myself, but I went to each of you privately and conferred upon him what I say is the greatest benefit, by trying to persuade him not to care for any of his belongings before caring that he himself should be as good and wise as possible, not to care for the city's possessions more than for the city itself, and to care for other things in the same way. (36c-d)

Whereas Socrates admonishes the Athenians for their lack of priorities in ranking the three types of goods in his self-reflective thought experiment, in the passage under consideration Socrates concludes his reflection on the worthiness of a life that was dedicated to the state of the soul. From this point in the dialogue to its conclusion, Socrates reflects on his present situation and the future of Athens.

Socrates' self-reflective thought experiment envisions the Athenians granting him the choice of being acquitted on the condition that he cease investigating and practicing philosophy or be put to death. The effectiveness of the thought experiment is made possible by specific features. The features of an effective thought experiment are efficiency, conceptualization, and refutation.

Socrates' thought experiment is efficient insofar as there is no possibility that he would cease practicing philosophy. It is designed to be a vivid reminder to the Athenians that philosophy requires sacrifices. His thought experiment also refutes the Athenians' way of life by showing them that ambition in the pursuit of reform is worthy of emulation if it is animated by moral ends, and that the virtuous soul, which is directed by wisdom, determines how bodily and external goods are put to practical use. Consequently, conceptual reform (i.e. appropriately prioritizing the three types of goods) and its effect on practical activity is made possible by the beneficial and useful function ascribed to virtue.[26]

V. Alcibiades Minor (141a-b1)

The last example of a thought experiment that emphasizes the connection between conceptual reform and practical activity is Socrates' discussion in *Alcibiades Minor*. Similar to the *Apology*, the thought experiment in *Alcibiades Minor* is in the form of counterfactual questions with imaginary answers. It is designed to make Alcibiades aware of the gulf between his ambition and his lack of knowledge. Such awareness motivates Alcibiades

[26] Xenophon's *Memorabilia* (2. 4-7) insists that Socrates' central characteristic is usefulness (esp. Xenophon's discussion of Socrates' approach to friendship). See Aristotle, *Rhet.* 1.9.1366a 36-8.

to reconceptualise his understanding of ambition by him coupling knowledge with his ambition. Consequently, Alcibiades comes to appreciate that there are practical limits to the pursuit of power.

In *Alcibiades Minor*, the young, ambitious Alcibiades encounters Socrates on the way to say prayers. Eager to be of help, Socrates insists that it would be better for Alcibiades not to pray for anything specific; but, instead, like the Spartans, pray for what is good.[27] Socrates' implicit claim is that human knowledge is fallible, we do not know what is best for us or what happiness is, and so it would be best not to pray for anything specific. He relates a cautionary tale to Alcibiades about being careful of what he asks for in prayer, because Alcibiades may "be praying for great evils when you think you are asking for great goods" (138b6-7). The tale is the story of Oedipus who inadvertently blurted out the prayer that his sons use arms to settle their inheritance, which eventually came to pass. Socrates then introduces the following counterfactual choices through a personified god:

> Suppose that the god to whom you are about to pray were to appear to you and ask you, before you began praying, whether you would be happy to be sole ruler of the city of Athens—or, if that seemed mean and tiny, were to offer you all the Greeks as well—or, if he saw that you regarded that too as insignificant unless the whole of Europe were included, were to promise you all of that plus simultaneous acknowledgement by the whole human race of the rule of Alcibiades son of Clinias. If that happened, I imagine, you will go home very happy and think you had come into possession of the greatest goods. (*Alcibiades Minor* 141a5-b1)[28]

Alcibiades agrees with what Socrates has just said, but Socrates queries Alcibiades on whether or not he would be willing to give up his life in exchange for the territory and all of Greece; or if he were to receive these great goods, if he would use them badly. Alcibiades quickly answers with a definitive no (141c).

---

[27] Socrates' assumption in offering the advice is that someone as ambitious as Alcibiades is bound to make specific requests of the gods. In fact, given Alcibiades' subsequent biography, Socrates' adversaries accused him of corrupting Alcibiades' ambition by inflaming his vanity. David Gribble (1999, 214-215) points out that there are two genres in the Alcibiades tradition. The *first* tradition, typified in the writings of Thucydides and Demosthenes, focuses primarily on Alcibiades' bios or 'way of life' and how it influenced his civic attitude. The *second* tradition, the Socratics, focuses primarily on Alcibiades as a moral agent shaped by his own choices as a young man. On Alcibiades' behaviour during the Peloponnesian War see Thucydides' *Peloponnesian War* (6.27-29) and Plutarch's *Life of Alcibiades*.

[28] Except where indicated, all translations from *Alcibiades Minor* are from Cooper 1997.

Alcibiades is very ambitious, as his responses to the thought experiment indicate, but it is not a reckless ambition. He is willing to forgo the territory and rulership if it means giving up his life, or if it means gaining these gifts but using them badly. The caveat Alcibiades makes regarding the counterfactual choices offered to him reveal that he recognizes that, without knowledge, there is no way to distinguish between precious and pernicious things. Of course, Socrates will be of some help in getting Alcibiades to distinguish between the two.

Due to the greater awareness that Alcibiades now has of factors that may thwart his quest to rule, such as a lack of a particular type of knowledge, Socrates is now in the position to stress the idea that ignorance of the best is what is bad in the use of possessions and actions. In other words, what is best, according to the scope of the craftsman's own skill, is being "mistaken about what is best for the state and for himself" (146a5-7). There are many examples that Socrates furnishes to reinforce the idea. For example, he argues that if one were in a state of good archers and flute-players, good athletes and craftsmen, each one in possession of a particular skill, but none of which had the knowledge of what is best, that state would be "a hotbed of dissention and lawlessness" (146b1-2). The idea that Socrates is expressing is that skilled practitioners need knowledge of utility in order to be of benefit to themselves and to the community; the knowledge of how to apply their craft in the real world. Similarly, practical knowledge serves as the basis of virtue when it is acknowledged that the Spartans only pray that they receive what is good and noble instead of praying for anything more (148c-150b3).

Conceptual Reform: Knowledge, Ambition and Irresponsibility

Might this idea of practical knowledge, functioning as the basis of virtue, resonate with Alcibiades? After all, the counterfactual choices offered to him revealed the length he is willing to go to be a great ruler. If he were to go before the Athenian people claiming to give advice on making war, like the orators claim to give advice on various topics, but was ignorant of what advice is best regarding war, might he fail to understand when to go, or for how long, and with whom? Alcibiades must not be irresponsible, and Socrates concludes his discussion by saying as much:

> For most people, then, it is an advantage neither to know nor to think they know anything, if they are going to do themselves more harm than good by rushing to do what they know or think they know. –Very true. –So you see it seems that I was quite right when I said that it looked as if other skills, if

> not combined with the knowledge of what is best, are more often than not harmful to their possessors. (146d1-e1)

This passage should be read in the context of the choices offered to Alcibiades in the thought experiment. Whereas, in the *Apology*, Socrates' thought experiment presents his own life to the Athenians as an opportunity for the city to reflect on its priorities, Socrates' thought experiment in *Alcibiades Minor* is designed to get Alcibiades to reflect upon the best or the good toward which his ambition should be used. His ambition to rule the city of Athens or all of Greece without having reflected upon the best or the good may lead him and the city to corruption. Socrates' responsible questioning, in the guise of a thought experiment, acknowledges the ambition of a youthful Alcibiades while also helping him see that understanding his ambition appropriately has good, practical consequences.

Consistent with the features contained in the thought experiments presented in the *Republic* and the *Apology*, the thought experiment in *Alcibiades Minor* is efficient insofar as it dramatizes the gulf between Alcibiades' ambition and his lack of knowledge. Having god offer Alcibiades choices that would gratify his ambition, Socrates dramatizes this gulf in a way that would not have been believable if he had offered Alcibiades the same imaginary choices. Socrates is not a god. It is the god-like perspective of the thought experiment that gets Alcibiades to reconceptualise his ambition and to submit to a conditional refutation of his complacency.

Conclusion

It should be clear from my discussion that the three Platonic dialogues I have looked at emphasize an aspect of thought experiments that is underemphasized in Kuhn's discussion. Kuhn's essay does not stress the effect of thought experiments on agent's actions. Kuhn's underemphasis would not be problematic if he did not call attention to the mainstream, contemporary view ignoring the joint implication of concepts and nature. His advice to proponents of the mainstream view is that thought experiments must teach the scientist or philosopher about his concepts and the world together. I have interpreted the latter point as suggesting that thought experiments sharpen an agent's conceptual apparatus, while at the same time making his actions in the world more precise and productive. I have argued that effective thought experiments in both the natural sciences and moral philosophy allow the experimenter to grasp the law-like

structures of nature or of morality. Unlike Kuhn, Plato emphasizes the point that grasping such law-like structures is inseparable from practical rationality.

Plato's thought experiments contain three features: efficiency, conceptualization, and refutation. Such features provide Plato the necessary emphasis lacking in Kuhn's standard discussion of thought experiments. Plato seems to take it for granted that thought experiments are effective precisely because of their practical effects. The Ring of Gyges and city for pigs (*Republic*), Socrates' self-reflective, counterfactual questioning during his trial (*Apology*), and Alcibiades being questioned by a personified god (*Alcibiades Minor*) are thought experiments grounded on the principle that if we could only straighten out our views, i.e. concepts, we would conduct ourselves correctly, and achieve the good for which we aim.

I hope my limited treatment of thought experiments in Plato's dialogues adds to a discussion that has the potential to illuminate other important aspects of thought experiments. Not only are Plato's ideas relevant in these sorts of discussions, the dialogues themselves provide a freshness that is often lacking in contemporary discussions of thought experiments.

Bibliography

Brown, James Robert. 2004. "Why Thought Experiments Transcend Empiricism". In *Contemporary Debates in Philosophy of Science*, edited by Christopher Hitchcock, 23-43. Oxford: Blackwell Publishing.

Burnyeat, Myles. 1997. "The Impiety of Socrates". *Ancient Philosophy*, Vol. 17: 1-12.

Burnyeat, Myles. 1980. "Virtue in Action". In *The Philosophy of Socrates: A Collection of Critical Essays*, edited by Gregory Vlastos, 209-211. Indiana: University of Notre Dame Press.

Burge, Tyler. 2007. "Two Thought Experiments Reviewed". In *Foundations of Mind*: 182-191. Oxford: Oxford University Press.

Carpenter, Michelle and Polansky, Ronald. 2002. "Variety of Socratic Elenchi". In *Does Socrates Have a Method?*, edited by Gary Scott, 89-100. University Park: The Pennsylvania State University Press.

Cohen, Martin. 2005. Wittgenstein's Beetle and other Classic Thought Experiments. Oxford: Blackwell Publishing.

Gendler, Tamar S. 2000. *Thought Experiment: On the Power and Limits of Imaginary Cases*. New York: Garland Publishing Company.

Gribble, David. 1999. *Alcibiades and Athens: A Study in Literary Presentation*. Oxford: Oxford University Press.

Glymour, Clark and Wimberly, Frank. 2007. "Actual Causes and Thought Experiments". In *Causation and Explanation*, edited by J. Campbell, M. O'Rourke, and H. Silverstein, 43-68. Cambridge: MIT Press.

Hanson, Norwood Russell. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.

Ierodiakonou, Katerina. 2005. "Ancient Thought Experiments: A First Approach". *Ancient Philosophy*, Vol. 25: 125-139.

Maurice, Andre. 2006. "Instances of Decision Theory in Plato's Alcibiades Major and Minor and in Xenophon's Memorabilia". *The Southern Journal of Philosophy*, Vol. 44: 365-380.

Kuhn, Thomas. 1977. "A Function for Thought Experiments". In *The Essential Tension*, 240-265. Chicago: Chicago University Press.

Norton, John D. 2004. "Why Thought Experiments do not Transcend Empiricism". In *Contemporary Debates in Philosophy of Science*, edited by Christopher Hitchcock, 44-66. Oxford: Blackwell Publishing.

Piaget, Jean. 1946. *Les notions de movement et devitesse chez l'enfant*. Paris: Presses Universitaires de France.

Plato. 1997. *Complete Works*. Edited by John M. Cooper and translated by G. M. A. Grube. Indianapolis: Hackett Publishing Company.

Plato. 2002. *Five Dialogues* (2nd Edition). Translated by G. M. A. Grube and revised by John M. Cooper. Indianapolis: Hackett Publishing Company.

Ramsey, Frank P. 1978. "Truth and Probability". In *Philosophical Papers*, edited by David H. Mellor, 60-100. Cambridge: Cambridge University Press.

Rescher, Nicholas. 1991. "Thought Experiments in Presocratic Philosophy". In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald J. Massey, 31-41. Maryland: Rowman & Lifflefield.

Rescher, Nicholas. 2005. *What If? Thought Experimentation in Philosophy.* New Brunswick: Transactions Publishers.

Sorensen, Roy. 1992. *Thought Experiments*. Oxford: Oxford University Press.

de Strycker, Emile and Slings, Simon R. 1994. *Plato's Apology of Socrates*. New York: Leiden Press.

Swirski, Peter. 2007. Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution and Game Theory. Abingdon-on-Thames: Routledge.

Vlastos, Gregory. 1991. *Socrates: Ironist and Moral Philosopher*. Ithaca: Cornell University Press.

CHAPTER TWO

FROM KALLIPOLIS TO UTOPIA: THE METAMORPHOSES OF IDEAL THEORY

NENAD MIŠČEVIĆ

Introduction: The Two Sorts of Platonopolis

The paper is about political thought-experiments[1] in the Platonic tradition.[2] Historically, it concentrates on Plato and the development of an ideal "Platonopolis" in the late antiquity,[3] and then makes a brief jump to their Renaissance descendants. Topically, its goal is connecting thought experimenting ("TE-ing" for short) and what is nowadays called "ideal theory". It points to famous examples in ancient philosophy of two important goals of TE-ing and the corresponding kinds of ideal theory, and to further historical philosophical developments.

So what is a philosophical TE? It is an exercise in philosophical imagination: in her introduction to the impressive collection on thought experiments in antiquity, Sophie Roux notes about TEs that

1. they are counterfactual,
2. they involve a concrete scenario, and
3. they have a well-determined cognitive intention. (Roux 2011, 19)

Tamar Szabo Gendler offers a similar characterization:

> To perform a thought experiment is to reason about a scenario /…/ which is imaginary /…/ with the aim of confirming or disconfirming some hypothesis

[1] I shall shorten "thought experiment" to "TE" and "political "thought experiment" to "PTE"."

[2] I wish to thank Boris Vezjak and other participants at the conference on TEs in antiquity.

[3] I am using the term from Dominic O'Meara's (2003) to cover all sorts of Platonic proposals concerning the ideal city.

> or theory. /…/ Contemplation of the scenario takes place with a specific purpose: the confirmation or disconfirmation of some hypothesis or theory. (Gendler 2004, 1155)

Applied to PTEs, her idea would, to my mind, yield the following specification:

(a) thought-experimental reasoning involves reasoning about a particular set of social and political circumstances, which may be specified in more or less detail,
(b) the thinker's mode of access to the scenario is via imagination rather than via observation, and
(c) the contemplation of the scenario takes place with a specific purpose: arriving at a judgment about some politically relevant theoretical proposal.

I shall argue, in the next section, that Plato's *Republic* is such a political TE. Then I want to connect this point to the idea of "ideal theory", following Rawls and his successors. Rawls has introduced the term "ideal theory", in his *Theory of Justice*, in a quite modest way, as the theory of the just arrangements that rely on the assumption of full compliance of the participants in the arrangement. He simply says at the beginning of his work that he shall, "for the most part /…/ examine the principles of justice that would regulate a well-ordered society" (Rawls 1999, 7-8), and then calls the resulting theory "ideal".[4] Of course, the biggest part of the discussion of this sense of "ideal theory" was dedicated to the relation with the non-ideal situation: what are we supposed to do if we know that citizens will not comply? In fact, Rawls' original suggestion is neutral in regard to the status

[4] Here is the relevant statement by Rawls:

> Thus I consider primarily what I call strict compliance as opposed to partial compliance theory (§§25, 39). The latter studies the principles that govern how we are to deal with injustice. It comprises such topics as the theory of punishment, the doctrine of just war, and the justification of the various ways of opposing unjust regimes, ranging from civil disobedience and conscientious objection to militant resistance and revolution. Also included here are questions of compensatory justice and of one form of institutional injustice against another. Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. (Rawls 1999, 8) For extended discussion see Simmons 2010.

of other characteristics of the just arrangement being discussed: we can imagine that it is a very demanding arrangement, or just a variant of existing ones. What makes it "ideal" in the first and weak sense is simple the assumption that participants comply with the rules of the arrangement.[5]

The full compliance meaning of "ideal theory" is too modest for our purpose of locating Platonic political philosophizing on the map, and we shall not discuss it further. Instead, we have to make the next step, as most discussants of the notion of ideal theory already have.

Let me note a suggestion by Laura Valentini from her excellent overview of the debate on ideal theory: "/…/ 'ideal theory' may be taken to mean 'utopian or idealistic theory', and 'nonideal theory' may be understood as 'realistic' theory" (Valentini, 2012, 654). It is obvious that ideal theory building typically is or involves thought experimenting, and we shall be illustrating this claim throughout the paper. The term has acquired new connotation in three decades of debates, and they shall be our topic in the fourth section of the paper.

Of course, the connection between ideal theorizing and Platonic construction has been noted in the literature (most clearly by Ypi, 2010; see the next section). I would stress the third relevant relatum, thought experimenting, which covers both components noted: ideal theorizing is, to a large extent, done through thought experimenting, and, as I shall argue, it is paradigmatically at work in Platonic construction.

So, in this paper, I would like to address the history of Platonic "state-building" heritage, by situating it in the framework developed by political theorizing. I start from the idea that this Platonic tradition exemplifies a lot of thought-experimenting, some of it purely epistemic (TE-ing in the strict sense), but most of it more motivational and advice providing—the typical job of famous historical Utopias. Then I place this quite obvious fact into the framework of ideal theorizing.

I. Plato's *Republic*: Thought-Experimenting and Ideal Theory

Plato's *Republic* marks the birth of what is nowadays called the "ideal theory". Its successors also introduced less theoretically and more motivationally oriented variants of ideal theorizing.[6] So, we shall work with

[5] The now standard source is Simmons (2010), but see also Hamlin and Stemplowska (2012).

[6] The characterization was made by Lea Ypi in her (2010) paper:

> We tend to think first and foremost of Rawls when differentiating between ideal theory and non-ideal circumstances. But of course the distinction

two sorts of Platonopolis, epistemic and motivational. This will be our topic here: first, to characterize the *Republic* and *The Laws* as belonging to a variant of ideal theory; second, to say something about the construction of ideal theory in the dialogical thought-experimenting; and third, to point to the birth of more motivationally oriented variants in late antiquity, in particular in the Byzantine traditions. We shall keep the term "Utopia" (with capital "U") for this motivational type of ideal theory.

The *Republic* is a strict PTE, with a central, epistemic purpose: to "watch a city coming to be in theory [or thought or discourse] /.../ [and] also see its justice coming to be, and its injustice as well" (Plato *Republic*, 369a)? Plato's famous proposal starts from a visual metaphor: it is easier to recognize justice in a polis than in the character of a single person, since, in a polis, it is written "with larger characters". But where does one find a just polis? Well, we can observe it in our *logos*, Socrates proposes.

> If we could watch a city coming to be in the *logos* <ἀρ' οὖν, εἰ γιγνομένην πόλιν θεασαίμεθα λόγῳ>, wouldn't we also see its justice coming to be, and its injustice as well? /.../ And when that process is completed, we can hope to find what we are looking for. (Plato *Republic*, 369a5-10)

Notoriously, *logos* means a lot of things, and has been translated in this sentence as "speech", "discourse", even "theory".[7] Moreover, Socrates elsewhere offers a story-oriented characterization of what is being done: ἐν μύθῳ μυθολογοῦντες (376d9).[8] What Socrates then really does is ask his interlocutors to imagine various arrangements and to judge them in terms of

---

> between ideal and non-ideal goes back at least to Plato. It emerges clearly in books 5 and 6 of *The Republic* (Plato, 1908) in which Socrates attempts to defend his ideal theory of the polity from one critique of Glaucon that he considers to be the 'greatest and heaviest wave' faced so far by his account. Glaucon's argument is that every previous reflection Socrates has offered on justice in an ideal polity relies on the assumption of full compliance and citizens' motivation to obey the polity's laws (Plato, *The Republic*, 471d). Yet these considerations seem to deflect the discussion from the real issue Socrates needs yet to address: 'is such an order of things possible, and how, if at all' (Plato, *The Republic*, 471e)? (Ypi 2010, 537-538)

[7] Grube (1992) has "in theory", Alan Bloom (1968) "in speech", Sydenham and Taylor (1906) "in our discourse". Jowett translates: "And if we imagine the State in process of creation, we shall see the justice and injustice of the State in process of creation also." Rouse (1999) omits the word logos and just has "see the city in the making". The old translation of Davies (1866) has "just imagine" and later in the text (on p. 369) "in idea".

[8] The suggested link with the narrative is worth exploring, but unfortunately we cannot do it here.

them being just. For example, Adeimantus will mention the topic of family. Socrates describes the arrangement involving commonality of children and then asks: is it just? Adeimantus gets the question right. He tries to imagine the arrangement and does it to his satisfaction. I would describe what he is doing as a tentative production or modelling. It may involve his experience of family life, his own, and of his acquaintances, some thinking about particular circumstance discussed in considering the previous proposals of Socrates, and, maybe, some other considerations.

At the end of his imagining, he asks himself the question about justice: all in all, would the community arrangement be just? It seems to him that it is.[9] He might feel some emotional warmth about it, and his immediate spontaneous answer in his thought is positive: "Yes, the community of children is just in this case." (This is a standard case of having a "moral intuition"). So he says what he thinks. Consequently, an arrangement has been built in *logos*, both in conversation and in Adeimantus' thought, and carefully observed. The inner observation has prompted the intuition-judgment that yes, the arrangement is just. Of course, a philosophically trained interlocutor or reader will immediately try varying and generalizing the point: can it be applied in any just polis? And he or she might end up with the general belief that children in every ideally just polis should be common to all citizens.

Since this is just one of many TEs making up the political core of the *Republic*, we might describe it as a "micro-TE", in comparison with the big aggregate piece of imagining and arguing that the book presents to its readers. Let us then consider our two issues. The inferentialist objection first. Could one arrive at the same conclusion just by using unproblematic premises, normative principles, and empirical observation and inference? A utilitarian might reply in the positive; here is her answer simplified to the extreme: consider the amount of enjoyment (and perhaps of suffering) that the arrangement brings to all concerned (or some other relevant consequence-value). If the amount is greater than in alternative arrangements, then the arrangement is just. A general value-consequentialist would need a way to assess the values relevant for the situation, and she might gladly appeal to TEs and intuitions. A deontologist would also be reserved about inferentialism. She would probably accept the need for immediate moral judgment, and for some kind of intuition, thus rejecting the inferentialist objection. So, the inferentialist is far from being the winner.

Consider now stages of TE and how they determine the ideal theory. Take the simple and famous example. Adeimantus is invited to imagine the cooperation between several people who work on agriculture, on production

[9] Some philosophers, e.g. Sosa (2007, 60), describe this kind of seeming as intuition.

of tools, on production of dresses, and the like. Socrates then notes that a worker will do his work better if he can

- "[370c] /…/ attend to it as his main affair, and not as a by-work." (I am using Shorey's translation here)
- Adeimantus agrees: "He must indeed."
- Socrates generalizes the morals: "The result, then, is that more things are produced, and better and more easily when one man performs one task according to his nature, at the right moment, and at leisure from other occupations." "By all means."
- "Then, Adeimantus, we need more than four citizens for the provision of the things we have mentioned. For the farmer, it appears, will not make his own plough if it is to be a good one."

We may assume that at *stage one,* the question is understood, one hopes correctly, by the interlocutor, in this case Adeimantus.

At *stage two* comes the tentative conscious production, his building of the "model" of the scenario at the conscious level.

Stage *three* might be an immediate, unconscious intuition: the worker will work better if he specializes in his task.

The *fourth stage* is more demanding. It concerns the production of the answer, involving the generation of intuition as to whether the arrangement is just or unjust. This probably involves reasoning at the unconscious level.

At *the fifth stage*, the thinker comes out with explicit intuition at the conscious level, usually geared to the particular example and having little generality.

Now, in many cases, the thinker will have to do some varying and generalizing (deploying both moral and rational competences) at the conscious and reflective level and, perhaps, at the unconscious one too. This is the *sixth* stage. Sometimes this process is called intuitive induction. (Chisholm, 1966).

Stage seven finally brings general belief at the reflective level. Three components are prominent in such a procedure: first, the aggregation of micro-TEs; second, the harmonization of the results of these micro-TEs; and finally, the judgment regarding their coherence with other moral intuitions one might have. In the example given, Adeimantus will confront further consequence of specialization, in connection with some facts about human nature (love of luxury) and come to a small but unified picture of the area to be understood. This philosophical unification can be described in terms of narrow reflective equilibrium.

One can call the harmonious unification of micro-TEs "the topical narrow reflective equilibrium", and the final narrow result "the general

narrow reflective equilibrium". The former, topical one, is geared towards the unity of narrative structure, plus the relevance and coherence between particular stages, the micro-TEs.

At *stage eight*, general knowledge of more empirical kind is brought into play: the important and difficult task of comparing the result with all we know about life and politics, both on a personal experiential level and from history, social and natural sciences, reaching a wide reflective equilibrium as the final result.

Now we see the role of reflective equilibrium in the construction of a small ideal picture: for instance, what the division of labour would look like in the good case. The particular micro-TEs offer the material, the reflective equilibrium ends with the ideal theory of the considered domain.[10]

II. Platonopolis: The Early Motivational Proposals

We can contrast two families of imaginative exercises, both present in the history of Platonism and in the general history of political philosophy. In the first the explicit goal is epistemic: find out what is justice, as stated in the *Republic*. The method: find/construct a system of arrangements, each of which appears just to the interlocutor, let him conclude that the system itself is just, and find what is characteristic of the system: this is then "the just". A further goal might be motivational: implementing the system as far as possible.

The second family, the one of historical political utopianism, characterized by the primacy of the motivational "Utopia", has at least two meanings. First, a very general, theoretical meaning: any presentation of a non-real, social-political scenario. This is a very popular meaning in the analytic debate: Schofield, Estlund, and the recent collection *Political Utopias: Contemporary Debates* by Vallier Weber (2017) come to mind.

The other meaning is historically more relevant: contrast pure PTE (epistemic) and utopia (motivational) the imaginative, thought-experimental exercise in thought concerning the just order cries for implementation. We shall use the lowercase "utopia" for the first meaning and the capital "Utopia" for the second one.

One might think that the difference between our proposed variants of ideal theorizing is superficial, and belongs to the mere pragmatics of the texts in question; some of them add, so to speak, the direction to the motivations of the reader. However, this does not hold. Motivational,

[10] For another interesting use of the notion of reflective equilibrium in the understanding of ideal theory, see Ben Laurence (Draft), "The Priority of Ideal Theory".

Utopian ideal theories are very frugal when speaking about more theoretical issues: what is justice, how we should characterize it and the like, on discussion of contrasting views and on argumentation, in particular of the more theoretical kind. The Utopia is described, and suggested to the reader as "the model" of the society, to use the term used by Francis Bacon in his *New Atlantis* (see the Preface), to be implemented by him (historically, it was the male reader who was targeted), if possible here and now. The philosophical contents are different, way beyond mere contrast in pragmatics.

So, where do we find such motivational proposals in the ancient times? Plato's direct Neoplatonic followers, Plotinus and Proclus, are obsessed with epistemic goals; indeed, they considerably expand the theo-cosmo-psycho-logical framework into which to fit the ideal state. So, when did the motivational writing start in the Hellenistic (or early Byzantine) period?

Let me try a conjecture: the clear motivational application of the Platonic ideal proposal is to be found in the times of Justinian, with authors giving advice to the emperor. Platonic ideas become candidates for motivating the ruler. Indeed, several Hellenistic "advices to the king" (see O'Meara 2003) can be connected in relation to utopian tradition, but are too practical-political to be taken to represent philosophically relevant utopias (see Bell 2009).

So, let me start with the "Dialogue On Political Science" from an anonymous author. Chapter Five is exactly what we need, a picture of The Ideal State. Bell offers a useful synopsis of the work, and I shall follow his picture and the picture offered by O'Meara. The main idea is that the ruler should imitate God; from it, the Anonymous derives concrete advice concerning matters like the selection of "High Priests" (or bishops); as well as the selection of the highest officials; next come "the nature, responsibilities and selection of the 'senate' of the 'optimates'" (Bell 2009, 143) and the office of emperor. The conclusion emphasizes the need for political education in rulers. O'Meara notes that, in general,

> [T]wo aspects of the divine, of God, were distinguished, knowledge, or perfect thought (θεωρìα), and providence, or care of what is lower. If the life of the divine has these two aspects, then the philosopher who is assimilated to the divine, or imitates it, will exhibit these two sorts of activity: theoretical activity, or knowledge, and providential activity, i.e. political rule. (O'Meara 2003, 178 ff.)[11]

[11] And here is more:

> If the anonymous dialogue describes a project comparable to the second-best city of Plato's Laws, then the criticism in the dialogue of Plat o's 'communism', the abolition in the *Republic* of private family life among the

Agapetus' *Advice to the Emperor Justinian* is quite idealistic. He notes that nothing gives man a better reputation than always "to want and what benefits mankind" (Bell 2009, 102, advice 6). He makes some propaganda for equality, and talks about the injustice of the differences between the rich and the poor. Inequality must be changed to equality, he concludes (Bell 2009, 105, advice 16). Bell notes about the demand "that Agapetus exploits it to stress not just giving to the poor, a Christian *topos* (or commonplace) but *taking from the rich* (cf. chs. 44, 51)" (Bell 2009, 105, fn. 27). And the *Advice* culminate in a somewhat Kantian suggestion: "Treat your servants as you pray that your Master will treat you" (Bell 2009, 108, advice 23).

Thus, in Justinian's era, the Platonic tradition turns to producing motivational, advice giving literature. It is fragmentary, it is not particularly deep, nor original, but it does accomplish what one would expect in Platonism: the production of motivational thought, of ideal theory directed to the potential force that can implement it. So much on the motivational writing in ancient Platonic tradition.

III. Utopia and Ideal Theory: Proposal for a Taxonomy

Let me pass to the next two tasks. One is to offer a wider, classificational framework for ideal theories, taking the epistemic/motivational framework as basic, and adding other distinctions. The wider framework will help us better understand the classical texts in contrast to later and contemporary alternatives. The other is to relate the Platonic project to the Utopian tradition, as it flourished in the Renaissance, and continued in modernity.

I shall begin with the second task, since it directly continues our engagement with Greek and Byzantine authors. The Renaissance utopian project explicitly continues the Platonic tradition. More's *Hexastichon*, announcing More's *Utopia*, famously speaks in the name of the Utopian island and states its ambition in the following terms:

elite (p. 22, 22–5) should be read, not as an attack on Plato himself, but as a rejection of this hallmark of the highest, divine, and indeed impossible city for humans in which all is held in common, a hallmark also rejected in the city of the Laws. I do not therefore think that we should conclude that the author of the dialogue, in rejecting this feature of the *Republic*, despite an obvious Neoplatonic inspiration, was not a Neoplatonist (O'Meara 2003, 182).

Plato's Republic now I claim
To match, or beat at its own game
For that was just a myth in prose
But what he wrote of I became
Eutopia is now my name. (More 2003, "Lines on the Island of Utopia")

However, our reading of the Platonic tradition should make us immediately sensitive to the issue of epistemic vs. motivational function of the philosophical proposal. Let us start with the full original title of the book: "*Libellus vere aureus, nec minus salutaris quam festivus*". *"Salutaris*" suggests a political role for the "booklet"; its point is to save us, present day readers, from social-political evils of our present day country. So, I would claim that More is recommending his ideal island as something "applicable to the demands of contemporary Christianity", as Colin Starnes, puts it (1990, 3). He does mention More's pessimistic assertion from the end of *Utopia*: "I readily admit that there are very many features in the Utopian commonwealth which it is easier for me to wish for in our countries than to have any hope seeing realized" (More 1965, 247).

Indeed, More does not present the island in the classical epistemic way reminiscent of Plato: More-Hythloday does not ask interlocutors to test each arrangement from the island, and then pronounce their verdict. The island is there, an ideal community, and the discussion is about chances for it to get implemented here and now.[12]

It seems that other Renaissance utopias belong to the same motivational category. Bacon's *New Atlantis* from 1627 is recommended (in its preface) to the reader by Bacon's chaplain, secretary, and literary executor W. Rawley, as a "model or mold" of both a college instituted for the interpreting of nature and the best state.[13]

As I mentioned, since the word "utopia" has several meanings, I shall call motivational ideal theory "Utopia" with a capital "U".

---

[12] For discussion of More's varying attitudes to the realizability of his Utopia, see Bevington 1961, 509; Skinner 1987; and. Starnes 1990. For further discussion of More's attitude, see also biography by Marius (1999).

[13] From preface, by W. Rawley:

> To the Reader
> This fable my Lord devised, to the end that he might exhibit therein a model or description of a college instituted for the interpreting of nature and the producing of great and marvellous works for the benefit of men, under the name of Salomon's House, or the College of the Six Days' Works /…/ His Lordship thought also in this present fable to have composed a frame of Laws, or of the best state or mould of a commonwealth" (Bacon 1992, Preface).

Petrić (Patrizzi), in his brief Utopia from 1554, *The Perfect City*, offers a description-advice on how to build the city that will guarantee happiness for its inhabitants; and again, there is no dialogue, no "testing of a hypothesis". In the dedication, he clearly uses a motivational metaphor of guidance: the book should provide powerful politicians (Vigerio and Girolamo della Rovere) with the description of the "path" that leads to a happy state. We can borrow his term and describe motivational ideal theories as guidance Utopias.

The most difficult candidate is Campanella with his 1662 *The City of the Sun*. Like in More and Petrić, the main narrator (the captain) gives his description to the grandmaster (*ospitalario*) as a travelogue; no discussion of the scenario presented, not a word. I would assume that the work is motivational, but that Campanella is silent about his motives given the very difficult situation with the Court of Inquisition he was in at the time of writing it.

However, there is no direct incitement for implementation, so there might be readers who will see *The City of the Sun* as a time-relativized ideal theory, but not as a Utopia in our sense of the word.

I shall assume that classical works are meant to propose ideal future arrangements, and that they also have a strong motivational component, that they are Utopias. If the reader disagrees about a particular author, say Campanella, she is free to stick with non-motivational reading.

We may conclude, with an apology for the extreme brevity, that the history of political utopianism is characterized by the primacy of the motivational: the Utopias are *libelli aurei salutares*, which explicitly means "to bring salvation to the society".[14]

This ends our discussion and illustrations of the epistemic/motivational distinction. Let me now introduce two more distinctions, within the genus of ideal theories and already partly present in the literature, to be combined with the epistemic/motivational one; I shall use them for the better understanding of the Platonic tradition. (I say "partly", since I shall interpret the first distinction in the way that has not been done in the literature).

The first distinction to be introduced is the one between time-relative and not time-relative theory. Let me start with a Platonic example. Plato proposes an ideal state in the *Republic*, but never relativizes it to time. Later utopias very often propose their ideal theories as theories of the end-state, or "end of history" as they put it somewhat politically. This is the distinction

[14] Note that the same text is sometimes read as more epistemic or more motivational, e.g. Locke's account of social contract is epistemic on traditional reading, but if we follow Ashcraft (1986), and read the Second Treatise as a political pamphlet, it will come out as motivational-applied.

I have in mind: the one between time-relative reading of an ideal theory, and a non-time-relative reading.

The classical utopian works are, of course, not always explicit about the ideal theory they propose. They typically put their object into the fictional present, or even past, like Plato playing with the phantasy of the Atlantis. Some of them also suggest that the proposed arrangement might be realizable in future, and that it is an ideal to be realized then and there.

Finally, one more distinction: some projects, like Plato's *Republic* and Rawls' *Theory of Justice* are quite idealizing: their clinging to the ideal is quite strong.[15] Others, like Plato's *Laws*, are less idealizing, their clinging to the ideal is moderate (one could introduce the third possibility, the weak clinging: e.g. one might argue that Aristotle's *Politics*, in particular book VII, is a bit ideal, just minimally so, and so are Cicero's *Laws*. We shall avoid this additional complication here, since it is not easy to say when the ideal is really, really weak). Let me quote Vallier and M. Weber:

> Political theory, from antiquity to the present, has been divided on the relationship between the requirements of justice and real-world barriers to meeting any such requirements, including limited human motivation, institutional limits, and scarce resources. Some theorists hold that a theory of justice should be utopian or idealistic—that the derivation of the correct principles of justice should not take such limits into account. Plato is a prime representative of this view in ancient thought. In contemporary political philosophy, G. A. Cohen is perhaps the standard bearer for idealism, as he holds that the correct principles of justice are completely 'fact-independent' in the sense that their justification does not depend on any real-world barriers to their realization. (Vallier and Weber 2017, 1)

Table 2-1

	EPISTEMIC	**MOTIVATIONAL**
A-TEMPORAL	**Strong** **Moderate**	**Strong or** **Moderate**
TEMPORAL	**Strong or** **Moderate**	**Strong or** **Moderate**

[15] See Valentini's formulation (2012, 656 ff.).

We can combine the two distinctions with our epistemic-motivational distinction, and characterize the theory of the *Republic* as non-temporal, strong epistemic ideal theory, and then look for end-state strong motivational ideal theory that would oppose it. We thus start looking for famous examples, authors that fit one or another pigeonhole in our schema.

Plato and early Neoplatonists clearly belong to the primarily epistemic category; motivation plays little role in their explicit discourse. They are not obsessed with history, and perfect state is not seen by them as the end-state institution, ending the history as we know it. As I just noted, Plato actually had a more idealizing project, the *Republic*, and the more realist one, *The Laws*. (Interestingly, Rawls belongs to the same category, with his views on justice and his movement from a more idealizing version in *The Theory of Justice* to the more realistic version in *Political Liberalism*. Early medieval thinkers, like Al Farabi, follow suit in both respects.

The more practically oriented advisors to Justinian are probably best seen as belonging to the motivational side: the wish for a state is for them something to be implemented by the actual emperor. Like Plato, their picture of the state is not time-relative.

Table 2-2

	EPISTEMIC	**MOTIVATIONAL**
A-TEMPORAL = TH. OF THE IDEAL	**Strong + Plato: *Republic***	**Moderate Justinian's advisors**
TEMPORAL	**Strong or Moderate**	**Strong or Moderate**

Let us now turn to famous Renaissance proposals of ideal theories. Remember that More points to possible usefulness of his theory, while Petrić-Patrizzi is much more explicit in his dedication: the short treatise should help the ruler, from the Della Rovere family, to guide his people to the mountain top at which happiness has constructed its paradise. His is thus clearly a case of what we have called "time-relativized ideal theory", our "Utopia" with capital a "U". Bacon himself is silent, but his editor, as we mentioned, proposes the Atlantis as a "model"; normally, it is read as an advice about possible ideal future arrangement(s). Campanella is silent, but

is so similar to other three works that we may take him as a motivational utopist.

Let me now add a few later names, mostly from nineteenth and twentieth century, with apology for brevity. First, as mentioned, we may add, on the side of a-temporal epistemic theory, the two proposals made by Rawls: a stronger, in his *Theory of justice,* and a more moderate, in his *Political Liberalism.* Also, to illustrate weak, non-demanding options, we can add Aristotle's proposal from *Politics VII*. On the a-temporal motivational side, let us add Bacon's *New Atlantis*; a plausible candidate from antiquity might be Cicero's *Laws*.

In the "Temporal" row, let me place two socialist utopians, both extremely sensitive to the historical context of their idealizing proposals. Charles Fourier has a surprisingly cold, objective way of presenting his strong, demanding ideas, of Phalanx and Harmony, with little propaganda, and an obvious wish to imitate the great scientists, physicists and biologists of his time. So, let me place him on the non-motivational, epistemic side.

Let me risk a bit, and add Cohen's *Why Not Socialism* on the epistemic side, in spite of his political engagement. On the very moderate end let me propose Fukuyama's picture of the end of history, practically realized in the actual reality.

The typical strongly motivationally oriented theorist is Robert Owen, with his vocal engagement for his New Lanark. The classical work, with the telling title, is *The Communist Manifesto* by Marx and Engels; very strong on motivational force, and very stingy on explanation, arguments, and other epistemic offerings for the reader. On the more contemporary side one can think of various social democratic projects and of cooperatives-proposing authors like probably Chomsky. So, let us place the authors we mentioned, into our classification scheme:

Table 2-3

	EPISTEMIC	MOTIVATIONAL
A-TEMPORAL = TH. OF THE IDEAL	**Strong:** Plato: *Republic* Al Farabi: *The virtuous state* Rawls: *Theory of justice* **Moderate:** Plato: *Laws* Rawls: *Political liberalism* **Weak:** Aristotle	**Strong:** Petrić: *Happy city* More: *Utopia* Campanella **Moderate:** Justinian's advisors **Weak:** Bacon: *New Atlantis* Cicero: *Laws-?*
TEMPORAL	**Strong:** Socialist utopias – Fourier Cohen: *Why not socialism* **Moderate:** Fukuyama	**Strong:** Socialist utopias-Owen Marx-Engels: *Communist manifesto* **Moderate:** Social democratic projects Chomsky-?

I am mentioning the authors from the second row in order to offer some examples to the reader; they don't belong to the main topic of our paper or the volume, so I do not discuss them at any length here, and shall leave it to the reader to decide whether they really fit the pigeonholes I placed them in.

At the beginning, I have mentioned one contrast not included in the table, Rawls' compliance-non-compliance distinction. Let me conclude by pointing to one more possible pigeonhole. We can think of "anti-ideal theory", a dystopian picture of reality to be avoided. The classics are easier to find in fiction, from Zamyatin and Orwell to Attwood, than in philosophy, but there are, for example, readings of Foucault on power that would present

his view as an anti-ideal theory. We thus end up, all in all, with eleven to twelve pigeonholes.

I hope the proposed classification schema works for all or most important political theories in history. But this remains to be tested in the future.

Conclusion

This paper has been addressing the history of the Platonic "state-building" heritage by situating it in the framework developed by later political theorizing. It starts from the idea that this Platonic tradition exemplifies a lot of thought-experimenting, some of it purely epistemic (TE-ing in the strict sense), but most of it more motivational and advice providing—the typical job of famous historical Utopias. Next, it places this quite obvious fact into the framework of ideal theorizing. PTEs are typically macro-TEs, combining a lot of micro TEs (ideally) into a coherent whole. The combination normally generates an ideal theory of some kind.

In the paper, we proposed a sketch of the partial change of the role of TE in the late antiquity, its replacement with purely motivational guidance-relative PTE, and a changed role of dialogue or its complete disappearance (Hellenism to Renaissance). We briefly pointed to interesting examples of Byzantine Neoplatonic advices to the emperor that can be read as burgeoning motivational TEs. In short, we have contrasted two families of imaginative exercises, both present in the Platonic tradition. In the first, the explicit goal is epistemic: find out what is justice. The method consists in finding/constructing a system of arrangements, each of which appears just to the interlocutor, in letting him conclude that the system itself is just, and in finding out what is characteristic of the system: this is then "the just". The further goal might be motivational: implementing the system as far as possible. This second family, the one of historical political utopianism, is characterized by the primacy of the motivational, by bringing the "guidance" or even "salvation" to one's society.

PTEs of both kinds have been central for political philosophy and theory, and the Platonic tradition is historically the first one illustrating this.

In the last section we combined our epistemic/motivational distinction with others, and proposed a table of kinds of ideal theories present in history, featuring ten or more kinds, in a way that is richer and more detailed than what is to be found in the literature.

We conjectured that, very probably, PTEs (like the *Republic*), and their motivational guidance-relatives organize the whole history of political thought, from antiquity to the present, so much more attention should be given to PTEs.

Bibliography

Ashcraft, Richard. 1986. *Revolutionary Politics and Locke's Two Treatises of Government*. Cambridge, Mass: Princeton University Press.

Bacon, Francis. 1992. *New Atlantis*. Montana: Kessinger Publishing.

Bell, Peter Neville. 2009. Three Political Voices from the Age of Justinian: Agapetus; Advice to the Emperor; Dialogue on Political Science; Paul the Silentiary; Description of Hagia Sophia. Liverpool: Liverpool University Press.

Bloom, Allan. 1968. *The Republic of Plato*. New York: HarperCollins Publishers.

Campanella, Tommaso. 1602. *The City of the Sun*. Frankfurt.

Chisholm, Roderic. 1966. *Theory of Knowledge*. New Jersey: Prentice Hall.

Bevington, David M. 1961. "The dialogue in *Utopia:* Two Sides to the Question". *Studies in Philology*, Vol. 58, No. 3: 496-509.

Davies, John Llewelyn and Vaughn, David James. 1886. The Republic of Plato, translated into English, with an introduction, analysis, and notes. Cambridge: Macmillan & Company.

Hamlin, Alan and Stemplowska, Zofia. 2012. "Theory, Ideal Theory and the Theory of Ideals". *Political Studies Review*, Vol. 10: 48–62

Gendler, Tamar Szabó. 2004. "Thought Experiments Rethought—and Reperceived". *Philosophy of Science*, Vol. 71, No. 5: 1152-1163.

Hunt, Lester H. 2015. "The Search for Utopia". In *Anarchy, State, and Utopia An Advanced Guide*, 200-224. Hoboken, New Yersey: Wiley-Blackwell.

Laurence, Ben. (Draft). "The Priority of Ideal Theory". Accesed December 20, 2020. [https://chicago.academia.edu/BenLaurence].

Ierodiakonou, K. and Roux, S., (Eds.). 2011. *Thought Experiments in Methodological and Historical Contexts*. Leiden: Brill.

Irwin, Terence. 1995. *Plato's Ethics*. Oxford: Oxford University Press.

Marius, Richard. 1999. *Thomas More: A Biography*. Cambridge, Mass: Harvard University Press.

Miller, Fred. D. 2009. "Aristotle on the Ideal Constitution". In G. *A Companion to Aristotle*, edited by G. Anagnostopoulos, 540-554. Oxford: Blackwell.

Miščević, N. 2013. "Political Thought Experiments from Plato to Rawls". In *Thought Experiments in Philosophy, Science, and the Arts*, edited by M. L. Frappier, Meynell and J. R. Brown, 191–206. London: Routledge.

Miščević, N. 2012. "Plato's Republic as a Political Thought Experiment". *Croatian Journal of Philosophy*, Vol. 35: 153-165.

Miščević, N. 2018. “Thought experiments in political philosophy”. In *The Routledge Companion to Thought Experiments,* edited by Stuart, M.T., Fehige, Y. and Brown, J.R., 153-170. London: Routledge.

More, Thomas. 1965. *Utopia*. In *The Yale Edition of the Complete Works of St Thomas More* (Vol. 4), edited by Edward Surtz, S.J. and J. H. Hexter. New Haven and London: Yale University Press.

O’Meara. Dominic. 2003. *Platonopolis: Platonic Political Philosophy in Late Antiquity*. Oxford: Clarendon Press.

Plato. 1969. *Plato in Twelve Volumes*. Vol. 5 and 6 translated by Paul Shorey. Cambridge: Harvard University Press.

Plato. 1997. *Complete Works*. Edited by John M. Cooper. Indianapolis/ Cambridge: Hackett Publishing Company.

Rawls, John. 1999. *A Theory of Justice*. Cambridge, Mass: Harvard University Press.

Roux, Sophie. 2011. “Introduction: The Emergence of the Notion of Thought Experiments”. In *Thought Experiments in Methodological and Historical Contexts*, edited by Katerina Ierodiakonou and Sophie Roux, 1-33. Leiden: Brill.

Schofield, Malcolm. 2006. Plato: Political Philosophy. Oxford: Oxford University Press.

Simmons, John A. 2010. “Ideal and Nonideal Theory”. Philosophy & Public Affairs, Vol. 38, No. 1: 5-36.

Skinner, Quentin. 1987. “Sir Thomas More’s Utopia and the Language of Renaissance Humanism”. *The Languages of Political Theory in Early-Modern Europe*, edited by A. Pagden, 123-158. Cambridge: Cambridge University Press.

Sosa, Ernest. 2007. *A Virtue Epistemology*. Oxford: Oxford Clarendon Press.

Starnes, Colin. 1990. *The New Republic: A Commentary on Book I of More’s* Utopia. Waterloo: Wilfrid Laurier University Press.

Sydenham, Floyer and Taylor, Thomas. 1906. *The Republic of Plato*. London: Methuen.

Vallier Kevin and Weber. Michael (Eds). 2017. *Political Utopias: Contemporary Debates*. Oxford: Oxford University Press.

Valentini, Laura. 2012. “Ideal vs Non-ideal Theory”. *Philosophy Compass*, Vol. 7, No. 9: 654–664.

CHAPTER THREE

THOUGHT EXPERIMENTING, HYPOTHETICAL THINKING AND IMAGINATION: SOME REMARKS ON PRESOCRATIC AND PLATONIC REASONING

BORIS VEZJAK

Is it methodologically possible, or correct and adequate, to ascribe the modern concept of thought experiment (TE), either "scientific", "philosophical" or "moral", to ancient philosophers and their theoretical thought? What conditions should be met in order to make such claims convincing? There is a plethora of different things that have, on some occasion, been called TEs, ranging from mathematical arguments, Presocratic reasoning, and Husserlian eidetic variation to Harvey's discovery of the circulation of the blood (Häggqvist 2009, 58).

Thought Experimenting "Ante Litteram"

Let us start with a general remark: what to do with modern philosophical or scientific conceptions when applied to the field of ancient philosophy? To put it quite coarsely, what is the approximate ancient Greek word that stands for "thought experiment"? Since the TE is such a central problem of analytic philosophy today, or even of philosophical task in a very general sense, it is tempting to suppose that typical ancient Greek authors, when dealing with problems of imagination and thinking, were primarily interested and involved in this topic as well.

Ierodiakonou (2018) believes that for ancient philosophers, TEs were used for three main purposes: to support, to attack, and to induce suspension of judgment about philosophical claims. Archytas of Tarentum's throwing a spear at the edge of the universe, the myth of Gyges in Plato's *Republic*, and Sextus Empiricus' "partless places" are three illustrative episodes; she

is not taking them as TEs, but refers to them as *paradeigmata*, or "examples". Since such supposition encounters obvious problems, we need to ask something else—a better starting point would be: are there any traces of the problem of TE (or any modern philosophical conception) in ancient philosophical texts? Let us try to pick up and explain another example among many: consciousness. If the Stagirite, for example, was interested in mental functions such as emotion and perception, was he primarily interested in this problem as well? How did Aristotle react or resolve the most difficult problems of consciousness? How do these types of concepts really differ from the examples of TEs?

In a rigorous sense, approaching some text in a described, conceptual manner would be dubious from a methodological point of view. As Tuominen (2009) noticed, such an approach implies a supposition that philosophical problems are constants that remain exactly the same throughout history. Consequently, can we cogently claim that TEs are such a constant? Since this supposition encounters obvious problems we need to ask something else. An even better starting point would be: are there any traces of the problems of TEs (or any modern philosophical conception) in ancient philosophical texts? When we ask this type of question, however, language and terms become problematic.

My possible objections against such a reading can be divided into the following structural scheme: (a) a general methodological objection: the status of modern conceptions or concepts in antiquity. What enables us to understand concepts like this as TEs or to recognize them within the framework of ancient philosophers?; (b) a motivational objection: were Presocratic hypothetical thinking or Plato's Ring of Gyges really meant to be presented as a TE? What was the real motivation behind such imaginative examples? Can we prove that they were reshaped as a TEs from previous or actual myths, or were they meant to be a simple, fictional story?; (c) a structural objection: are the elements of Presocratic and Plato's reasoning supporting or not supporting the idea of reading it as a TE?; (d) an interpretative objection: how are Presocratic or Plato's stories read by most scholars and how do they differ from readings of TEs?

Some indirect evidence for this can be provided: there is perhaps no Greek term that could be decidedly translated as "thought experiment". Taking philosophical concepts as something constant is surely a matter of different methodological standpoints or even traditions—analytic philosophy was always accused of a more ahistorical or anti-hermeneutical stand; for analytic thinkers, the conceptual scheme and truth were always atemporal, while continental philosophy always treated them within temporal or historical contexts. Is accessing or treating concepts "ante litteram" in itself

then only a version of a historical and conceptual reduction? However, following Tuominen (2009), even in the case where we find one of these terms, we are not allowed, without further consideration, to jump to the conclusion. What criteria need to be fulfilled to enable us in understanding these terms as referring to the feeling aspects of experiences? Evidently we lack some further explanation to show this. Again: when we ask this type of question, language and terms become problematic.

Let us call such a problem in methodology the "ante litteram/avant la letter" question in philosophy, meaning "(long) before the term was invented", or sometimes "before a discipline (i.e. anthropology) was formalized". "Ante litteram" concepts obviously refer to those invented before a term was coined. Let us take into consideration the following example: Herodotus gave an anthropological description "avant-la-lettre" of the European character, because anthropology became a scientific field of study in the 18th century. Some basic evidence for this can be provided: there is perhaps no Greek term that could be unquestionably translated as "thought experiment". Similarly, perhaps there is not even a single Greek philosopher conceptualizing perceptions and consciousness; the word for perception was *aisthesis*, intellectual thought or apprehension (*nous*), emotion or feeling (*pathos*), and many other mental processes. As Touminen (2009) stresses, maybe no Greek equivalent exists to refer to the general idea that we are conscious beings, which would more specifically mean that it is difficult to find a word that would capture the idea that experiences have an inner feeling or quality to them. Aristotle, for example, was interested in mental functions such as emotion and perception, was he primarily interested in this problem as well? How did he react or resolve the most difficult problems of consciousness? How do these types of concepts really differ from the example of TE?

Like in the case of consciousness, some ancient authors indeed employed terms like *suneidêsis* and *sunaisthêsis*; such concept could be taken to refer to the inner quality of these processes. However, again following Tuominen (2009), even if in this case we find one of these terms, we are not allowed, without further consideration, to jump to the conclusion that the problem of consciousness is at issue. What criteria need to be fulfilled to enable us in taking these terms as referring to the feeling aspect of experiences? Evidently we lack some further explanation to show this. I would like to advocate for more scepticism in the discussion of thought experiments. Obviously, we need to proceed in a way that first clearly defines TE and then checks whether individual cases and contexts meet the initial definition. Such line of interpreting introduces some further complicated issues: which philosophical stories in the early history of

philosophy count as TEs and which count as something else? Is there any prehistory of thought experiments that became a standard procedure in early modern science and can be traced back into ancient and medieval philosophies in different contexts?

Some authors, like Rescher (1991), claim that even some of the Presocratic philosophers of nature provide us with a bunch of good examples of TEs. On the other hand, authors like Bealer (1998) claim exactly the opposite and seriously doubt the possibility of using the term as referring to the ancient practice of appealing to intuition. If many philosophers take TEs as playing an essential role in analytic philosophy, Bealer is inclined to talk of rational intuitions: "To call them thought experiments is, not only to invite confusion about philosophical method, but to destroy the utility of a once useful term" (Bealer 1998, 208). This seems to be even more suitable in the cases of ethical and non-scientific TEs. Obviously, the given contrast between expanding the field of TEs into Presocratic philosophy (like Rescher) or totally reducing its practice to intuitions on the other hand clearly shows a variety of different readings and methodological approaches.

Ierodiakonou (2011, 37) speaks of the first recorded thought experiment in antiquity, attributed to the Pythagorean Archytas of Tarentum in the first half of the 4th century, about a man standing at the edge of the universe and extending his hand or his stick—trying to prove that the universe is infinite. What we need is more conceptual work done in examining the vocabulary and the notions in the works of ancient philosophers. Some philosophers suppose that there are cases where a TE is used in a non-metaphorical way.

Thought Experiments in Presocratics

Some authors, like Rescher (2005a), clearly believe that TE describes something nothing less than an attempt to draw instruction from a process of hypothetical reasoning that proceeds by eliciting the consequences of a hypothesis. That would imply that TE is really some sort of a general logical procedure, it is a sort of reasoning from a supposition that is not (or not yet) accepted as true—and perhaps is even known to be false—but is assumed provisionally in the interests of making a point or establishing a conclusion.

For him, the process of TE in philosophy imitates the one in natural science and "is as old as the subject itself" (Rescher 2005a, 48). Moreover, it was already a prominent instrumentality in the thought of the Presocratic nature-philosophers of ancient Greece, and therefore constitutes a part of history of modes of argumentation and reasoning. Let me briefly sum up

Rescher's arguments for thinking that Presocratics were already practicing some sort of thought experimentation.

Explanatory conjectures: first, Rescher's method is through explanatory conjectures (Rescher 2005a, 48-50). This means that thought experimentation is explanatory in character through a line of reasoning, such as already used by Thales of Miletus (born c. 620 B.C.), the first of these Presocratic nature philosophers. As we learn from Aristotle and his followers, Thales thought that the flooding of the Nile occurs because "the Elesian winds, blowing straight on to Egypt, raise up the mass of the Nile's water through cutting off its outflow by the swelling of the sea coming against it." (*Aetius* IV, 1, a; Kirk and Raven 1957, 77.). He also thought "that the earth floats on the water, and that it stays in place though floating like a log or some other such thing /…/" (Aristotle, *De Caelo*, B13, 294a28; Kirk and Raven 1957, 87.) And he declared the psyche to be made of water, apparently being persuaded by considering seeds, which are all moist (Aristotle, *De Anima*, A 2, 405 bl; Kirk and Raven 1957, 90.)

This would lead us to the use of TE in the context of explanatory conjectures; answers to the following pattern, where we assume that it is to be shown that P is the case (where it has not yet been established whether P or not-P). Then we assume—as TE—that P is the case (which is not inherently implausible). Then we explain Q on the basis of this assumption, where Q is something patently true that we could not readily explain otherwise. Hence we maintain that P. For example: consider the case of the psyche and the principle of life, regarded as that which enables living things to be living, and which all living things thus have to have in common. Here is the situation: we have to show that the psyche is made of water. Assuming this to be the case, we suppose that the psyche is made of water, which naturally explains why all seeds both have moistness in them and need water to develop. Therefore, we are justified in claiming that the psyche is made of water. Rescher thinks that such positive and productive use of TE for explanatory purposes in contexts of what ultimately came to be known as "hypothetico-deductive" reasoning represents their oldest and no doubt most familiar employment.

Negatively demonstrative reasoning: the second method suggests the most common uses of explanatory TEs proceed by way of analogy, done through negatively demonstrative employment of thought experiments. Here is one of his examples: Anaximander of Miletus maintained that Earth is at the world's centre and thus Aristotle attributes to him the following reasoning: "[T]hus if the earth now stays in place through the operation of a force, it too comes together at the center by being carried there because of the vortex" (*De Caelo*, B 13, 295a7; Kirk and Raven 1957, 127).

Rescher (2005a, 50-52) thinks this involves the reasoning of how in vortices objects tend to the centre; then Anaximander supposed the world to be vortex-like, and clearly this would explain that a large solid object like the earth would come to be positioned at the centre. The thesis supported by means of the analogy is clearly being argued for by the same method of explanatory thought-experimentation in Thales, but in Anaximander there is a negatively demonstrative use of thought experimentation, rather different from such explanatory employment. Thus let us consider the following justification for Anaximander's contention, that "the earth stays aloft, held up by nothing, but remaining in place on account of its similar distance from all things" (Hippolytus, *Refutatio haeresium*, 1, 6, 3. Kirk and Raven 1957, 134) And, "It stays still because of its equilibrium. For it behooves that which is established at the center, and is equally related to the extremes, not to be borne one whit more either up or down or to the sides." (Aristotle, *De Caelo* B13, 295b10; Kirk and Raven 1957, 134.)

This means that if the earth were not at the centre, then it would eventually succumb to a tendency to move further in one direction or another, and so would not have a stable, fixed and firm position at all. Rescher then suggests that, in such use of TE, the approach is negatively demonstrative reasoning: we would like to show that P (where we do not yet know whether P or not-P). Then we assume—as TE—that not-P, and deducing Q on the basis of this assumption where Q is some patently false thesis. Hence we maintain that P. In his view, such negatively demonstrative employment of thought experiments is characterized by its refutatory use. It is based on the well-known principle of indirect or "apagogical" reasoning that concludes negatively where a correlative positivity entails a false consequence, where he mentions that this reasoning is aptly characterized in Alexander Gottlieb Baumgarten's *Logica* as "demonstratio falsitatis alicuius propositionis ex sequentibus ex illa falsis" (sect. 691).

Reductio ad absurdum: the third method, found in the school of Pythagoras of Samos, is the negatively probative mode of hypothetical reasoning, transmuted into a formal mathematical method of proof—the mode of demonstration that has come to be known as reductio ad absurdum argumentation (Rescher 2005a, 53-54). We demonstrate that P, then we assume—as TE—that not-P, deducing an outright contradiction from this assumption (this is generally effected by deducing P itself). At the end we establish P.

The illustration of it is famous proof of the incommensurability of the diagonal of a square with its sides, which is the standard way of establishing the irrationality of the square root of two—one assumes the contrary as a working hypothesis and derives a contradiction. Rescher concludes that a

reductio represents a further development in the use of thought experiments—the transmutation of the negativity-productive mode of thought experimentation into a formal method of mathematical proof.

Sceptical thought experimentation: the fourth method was inaugurated by Xenophanes in a style of more sceptical use of thought experimentation (Rescher 2005a, 54-57). A classical instance of this sort of reasoning could be as following:

"But if cattle and horses or lions had hands, or could draw with their hands and do the works that men can do, then horses would draw the forms of the gods like horses, and cattle like cattle, and they would make their bodies such as they each had themselves." (Kirk and Raven 1957, 169, fragment 15; Clement, Stromata, v, 109, 3.)

For Rescher, such style of reasoning depicts some sort of sceptical thought experimentation in the following form: things being as they are, we are inclined to accept that P must be true. But then we suppose, by way of a TE, that our situation would be appropriately different. After that we would not accept P at all, but rather something else that is incompatible with P. Hence we are not really warranted in our categorical acceptance of P, seeing that, after all, this is merely a contingent aspect of our particular, potentially variable situation.

Or let us take another Rescher's example: "If god had not made yellow honey, men would consider figs far sweeter" (Kirk and Raven 1957, 180, fragment 38.). The pattern we get is that first we have things being as they are; honey is "the sweetest thing in the world"—the very epitome of sweetness. Then we suppose that honey would not exist. In this case, figs would be the sweetest thing we know of, so they would be the epitome of sweetness. Hence we should not maintain that honey is actually the epitome of sweetness; it merely happens to be the sweetest thing we happen to know.

Such argumentation is also clearly instantiated in his wisdom on how no man knows, or ever will know, the truth about the gods (Frag. 34; Kirk and Raven 1957, 179.) The procedure shows, according to Rescher, how Xenophanes relied on thought experiments to establish the relativity of human knowledge and scepticism about it.

Analogical thought experimentation: the fifth method (Rescher 2005a, 57-58) is exemplified via the analogical use of TE described in the following paradigm: "They vainly purify themselves by defiling themselves with blood, just as if one who had stepped into the mud were to wash with mud. Anyone who saw him doing this would deem him mad." (Frag. 5/129 and 130; Burnet 1920, 145; Kirk and Raven 1957, 211.)

The idea is that we could call someone who tries to remove mud with mud crazy; what then of those who try to remove blood with blood; will we

not have to call them crazy too? asks Rescher. This analogy-exploiting, critical use of TE is, in his view, clearly something quite different from its explanatory use as exemplified in Thales. It has the following form: suppose someone did X. Then one would say that he is F (mad, bad, or the like). But doing Y is just like doing X in the F-relevant regards. Therefore, one should also say that someone who does Y is F.

Value dominance argumentation: let us call this the sixth method—Heraclitus uses such sort of reasoning repeatedly to argue for the mutual dependence of opposites, "It is not good for men to get all they wish to get. It is sickness that makes health pleasant; evil, good; hunger, satiety; weariness, rest" (Frag. 111/104; Burnet 1920, 140.). This style of argumentation (Rescher 2005a, 58-60) is evidently tailor-made for a thinker who held that the mutual interdependence of opposites establishes the co-equal importance of the conceptions at issue: "Men do not know how what is at variance agrees with itself. It is an attunement of opposite tensions, like that of the bow and the lyre." (Frag. 51/45; Burnet 1920, 136.)

Here thought experimentation, along the indicated lines, can be used to show that in removing the tension we also destroy the very object that is at stake. As these considerations show, Heraclitus was a devoted practitioner of thought experimentation, believes Rescher, given to extracting far-reaching conclusions from fact-contravening hypotheses. The form of the argumentation is the following: Assume—by way of a TE—that X did not exist. Establish that, in this event, we could not even form the conception of Y, seeing that X and Y are correlative concepts (hot/cold, cause/effect, etc.). Conclude that therefore Y's place in the overall scheme of things cannot be less important or valuable than X's.

What he believes is that thought experimentation employed by Greek Presocratic philosophers of nature was a salient methodological device for developing their ideas. For the Presocratics, conjecture was not a "creative activity pursued for its own speculative interest, but an instrumentality for the investigation of the realm of truth and reality" (Rescher 2005, 60), however he does not equate such pioneering practice with "real experimentation".

Rescher was not the first one to use TE in its negative or refutatory use. James Robert Brown (1991, 34) distinguished three different kinds of thought experiments: constructive, destructive, and Platonic. The latter are those that are at the same time both constructive and destructive. His example for constructive thought experiments is Newton's rotating bucket thought experiment, which serves to provide evidence in support of a theory; in this case, Newton's theory of absolute space and time. For destructive thought experiments, Brown offers the case of Schrodinger's cat

as an example: destructive TEs serve to overthrow a proposed theory. In this case, Schrodinger's target was the Copenhagen interpretation of quantum mechanics. Irvine (1991) also explains that TEs are a key part of science. They are in the same realm as physical experiments; they require all assumptions to be supported by empirical evidence. The context must be believable, and it must provide useful answers to complex questions. Furthermore, TE must have the potential to be falsified. Nevertheless, he locates the origin of TE in the Presocratic Greek epoch, largely identifying thought experimentation with hypothetico- deductive reasoning like Rescher: it was the Presocratics who introduced the use of thought experiments in their reasoning about nature and, in doing so, it was they who introduced a versatile, efficient instrument that would prove to be essential for later development of the sciences (Irvine 1991, 153).

However, and unexpectedly, very few people rejected Rescher's views on TE. In other words, it is not clear at all why the procedures he uses should be understood as basic examples of thought experimentation rather than fundamental forms of logical reasoning. To consider some of the logical procedures as TEs, however, seems to be an unexpected statement or even a severe conceptual slip. Ierodiakonou and Roux (2011, 33) only mention his inspiration from a counterfactual thought experiment, and that he treats TEs as argumentative procedures resembling tests of consistency, which invite the experimenter to seek the weakest link in her body of beliefs. The strong counterfactual dimension of TE is, in reality, counterfactual reasoning—to be distinguished from counterfactual declarations, which pose totally different problems and raise totally different questions (Ierodiakonou and Roux 2011, 168).

From Presocratics to Plato's Gyges

The invoked myth of the Ring of Gyges (Plato, *The Republic*, 359a–360d) is believed to be one of the first examples of a thought experiment (TE) in the history of Western Philosophy. Some scholars maintain that it is established through means of a real mental experiment: as in many other cases, it was set up as a thought experiment and not only used metaphorically. Can we also be sceptical about this famous example as a real and perfect TE, even though it is widely thought to be convincing and very close to the basic modern definition of TE?

In a famous discussion in Plato's *Republic*, Glaucon maintains that those who practice justice do it unwillingly and because they lack the power to do injustice. Consequently, the only reason that anyone acts justly is out of fear that they will be caught or maybe punished. What Glaucon argues for with

Socrates is the commonplace belief that justice is not for its own sake, rather it is preferred because of its consequences: people who practice it do so reluctantly, regarding it as something necessary and not as a good. By introducing two hypothetical examples, the legend of the Ring of Gyges and the choice of lives, an attempt is made to demonstrate that as long as one's reputation is not in danger, one will have no desire to act morally.

In order to understand the motivational objection behind Plato's myth, we need to propose the most plausible explanation about why Plato decided to introduce it, or, even better, explain why has he probably changed the content of the original story. Of course, the main line of motivation is to provide a firm answer to the question of human morality: under which circumstances and why do we act morally? The Gyges story plays a major role in setting the main argument, and is maybe one of the key points in *The Republic*. The challenge is put forward by Glaucon and Adeimantus at the beginning of Book 2, and only comes to an end in Book 10, when Glaucon is asked by Socrates to give to justice back its usual reputation and rewards (Plato, *The Republic*, 612a8–e1).

Glaucon then presents two metaphors or supposed TEs, where injustice appears to prevail, and asks Socrates to defend why one should act justly in each of these scenarios (Vezjak 2017). The first theory presented is that of the "Ring of Gyges", the shepherd of the king of Lydia. We are told the rich, mythical tale about Gyges, discovering an underground cavern full of wonders after the earthquake. After descending into the cavern, he finds, on top of the hand of a large corpse, lying inside a hollow bronze horse, a magical ring capable of making its owner invisible. He then uses this ring to enter the king's palace, to seduce the queen, and together with her kill the king and take over the throne. According to the well-known story, this ring had some special powers of invisibility. "If he turned the setting inward, he became invisible; if he turned it outward, he became visible again" (Plato, *The Republic*, 360A). If we were in possession of such a mythical ring, nobody would act justly: so Glaucon claims that even the most just man would behave unjustly.

Plato's story thus proves that people are only just because they are afraid of punishment for injustice; therefore, no one is just because of justice itself. Additionally, Glaucon also obviously tries to demonstrate that not only is being unjust better than being just, but also that it is rational to choose the first possibility. He is suggesting that we already, in some sense, know how a just person would act under conditions of impunity; we just do not realize it because we have not organized the information in a way that makes its implications apparent to us. Such an argument is supporting the conventional belief that justice is only worth practicing because of its

consequences, not for its own sake – this is why a true defence of justice can only be advanced without referring to its consequences.

The Gyges story is introduced to support the second part of Glaucon's argument and intended to confirm some widespread beliefs; however, it also appeals to and depends on such beliefs. The just man would act exactly the same as the unjust man if he could only avoid detection and punishment (Plato, *The Republic*, 358b8–c6). The very same explanation is built into the narrative and the subsequent analysis: upon realizing the extant of the power he has gained because of the ring, Gyges immediately takes steps to usurp the king, the final conclusion being that no man with a similar power, no matter how virtuous he might be, would be able to resist engaging in all kinds of injustices. It is of an utmost importance to mention, in order to know exactly what Plato's motivation of introducing the story was, that his myth might only be a version of a previous myth. It is also found in Herodotus, the reputable Greek historian known as the father of history, and some other authors (in Herodotus' Histories, 1.8-14).

According to Miščević (2012), the story about the Ring of Gyges serves as one of the finest first examples of TEs in the history of European philosophy and proceeds as a typical "imaginative exercise", which is done in the armchair without physical experimenting or observation, and is aimed at discovering the truth (or something akin to truth) about some given issue. As he remarked in his paper on Plato's *Republic* as a political thought experiment (2012), they typically concern unreal situations, although some of the scenarios might have been actual, or might be made actual. The same view is shared, for instance, by Gendler (1996). She also speaks about a technique, a basis of the scientific method, used by Plato. Obviously, we are dealing with the conception of the story of the Ring of Gyges, which is affirmative in taking Plato's metaphor as the example of an ethical TE. The experiment Glaucon proposes in *The Republic*, writes Gendler, is not to be actually conducted. Rather, she suggests, we can see his point "most clearly /.../ if in our thoughts we grant to a just and an unjust person the freedom to do whatever they like" (Plato *The Republic*, 359bc). Glaucon is suggesting that we already, in some sense, know how a just person would act under conditions of impunity; we just do not realize it because we have not organized the information in a way that renders its implications apparent to us.

I already challenged the possibility of understanding the Ring of Gyges as a fine example of a TE, i.e. in what way is Plato really offering us something that would fit the common description of a TE (Vezjak 2017). My scepticism on this issue can be expressed through very different approaches, and I will try to describe a few basic sketches of it: was the

story meant to be a TE? Is such an explanation too quick and insufficient and can we demonstrate that Plato's myth might be better if interpreted differently and not necessarily as an early example of a TE? I will claim that there is a difference between a hypothetical assumption or making an analogy as proposed by Plato, and thought experiments in philosophy as used today. The same goes for the imaginary scenarios that Plato advances in *The Republic* or in *Meno*, in the famous example on geometry. Obviously, there are some relations to some modern TEs, but at the very least we cannot take these imaginary scenarios as the only possible way of reading Plato's mythical suggestion.

Many authors, like Miščević (2013) and Gendler (1996), do not approach the issue of Plato's possible usage of a TE in the case of the story of Gyges as something problematic; they take it for granted. When it comes to the description of an ideal polis in *The Republic*, Miščević advocates, as in the case of Rawls's *Theory of Justice*, that Plato's dialogue as a whole is a good example of a political TE, and describes it as an "imagined political arrangement" (Miščević 2013). Usually, we face different approaches when trying to apply TEs within a theoretical context. Since TEs have variously important roles in the very heterogeneous disciplines like mathematics, physics, literature, and philosophy, it should be noted that it is reasonable to treat them separately from, e.g. metaphors—they simply follow different goals. Sometimes this notion is taken in a very broad sense. Can we really claim that Plato's "Cave", Descartes's "malin genie", Hobbes' vision of the primal state of nature, or Kant's proposal of the categorical imperative are all examples of TEs? If we expect that a TE will serve its purpose, some kind of knowledge is supposed to be derived from it.

In many instances, the main goal of a TE, especially in the field of ethics, can hardly be obtained: we would like to analyse what happens when applying our theoretical ideas in order to test hypothetical findings, and at the end accept some theory. When Descartes, for example, imagined his "malin genie", we accepted his hypothetical scenario in order to reveal whether our knowledge from the beginning was reliable and trustworthy. At the end, we ended up with some new knowledge and insight. Dennett (1984) has drawn attention to the fact that many TEs are used in arguments that rely on making an appeal to intuition. In such a case, TEs are nothing more than "intuition pumps" – they are not supposed to provide strict arguments that prove conclusions from premises, their point is rather "to entrain a family of imaginative reflections in the reader that ultimately yields not a formal conclusion, but a dictate of 'intuition'". (Dennet 1984, 12)

This is surely a very vague description of its functioning. But even intuitions are always a matter of knowing. A possible objection would be

ascribing something else to vivid or metaphorical ideas (since they are not necessarily given as hypothetical scenarios), which would depict the world in such a way that it will be informatively processed into new knowledge for us. Frequently, it is only a metaphorical or literary description of something that we already know. In this sense, some common philosophical examples fulfil such a criterion and some do not; TEs tend to be, in such cases, more similar to some metaphorical or allegorical content.

The question is, of course, whether we can legitimately demand that ethical TEs operate in the same way as typical scientific TEs. A more likely motivation behind Plato's story of Gyges is that he was trying to confront us only with some kind of an analogy and imagery: yet there are several incoherencies in Plato's method and methodology of the middle dialogues, the Ring of Gyges being only one of many philosophical images in *The Republic*. Not many of them can be treated as a TE. If we accept the possibility of Plato's intention to reshape the previous or actual myth in order to enrich it and make it more than a simple fictional story, then the same can be done for all other cases of metaphors, analogies, and stories in *The Republic*. In each case, an imaginary situation is constructed, then somehow manipulated or analysed. In a series of scenes from *The Republic*, for instance, to show that a state is or can be perfectly just; in the case of a ship filled with sailors fighting for control; in the case of a cave containing prisoners staring at shadows on a wall; the divided line itself; etc.

We can agree that the Ring is more than a metaphor—the purpose of it is not only in illustrating an idea, but in providing some new understanding or insight. Comparing it to Plato's cave allegory as an example of a metaphorical device, the Ring is somewhat less rigid. The TE is intended to test what happens if we apply our theoretical ideas—in *The Republic* those ideas are about justice and human morality. After testing different hypothetical results, we are ready to accept such ideas, but only if we can accept them.

Let us mention the structural objection and the puzzle of imaginability: are the elements of Presocratic reasoning or Plato's Gyges myth supporting or not supporting the idea of reading it as a TE? Sometimes they depict an imaginary situation or an event in order to test our common sense beliefs and test whether our common sense beliefs hold up. If they do not, they should possibly be re-examined. According to Brown and Fehige (2016), TEs are

> /…/ devices of the imagination used to investigate the nature of things /.../ They should also be distinguished from counterfactual reasoning in general, as they seem to require an experimental element, which seems to explains

> the impression that something is experienced in a thought experiment. (Brown and Fehige 2016, Introduction)

Can we agree that such TEs fulfil all the elements required in the above mentioned definition? Brown's description seems to be rich and somehow empty at the same time—to say that there is "an experimental element" in TEs and that there is something experimental about it sounds more like a needless tautology without providing a good criterion. Gendler (1996) also thinks there is "a structure" in Glaucon's line of thought. In the second part of the story, the setting of the ring inward or outward makes the wearer invisible or visible. But if there were two such rings, one worn by a just and the other by an unjust person, both would follow the same path, their actions would be in no way different.

"Whenever either person thinks he can do injustice with impunity, he does it" (Plato, *The Republic*, 359c–360c). Thus Glaucon's TE fits into the tripartite structure, claims Gendler: he begins by presenting (1) an imaginary scenario, i.e. two rings are described and presupposed to have supernatural properties—making someone visible or invisible; (2) there is reasoning within the context of this imaginary scenario, namely, Glaucon contends that armed with the freedom to take whatever he wants with impunity, the just person acts exactly like the unjust person; and (3) an applied result of his reasoning to the actual world in the sense of: "Whenever either person thinks he can do injustice with impunity, he does it."

A TE is, for her, functioning like an ordinary experiment: it takes information already (in some sense) available to us through ordinary experience and allows us to organize it in such a way that its implications become apparent. Nevertheless, Gendler fails in showing why these structural elements adequately describe something as a TE. Also, in the case of the Ring of Gyges, there is no distinction between the method of thought experimenting and the result of such a method. Miščević (2013) believes that Plato started one of the two dominant political TEs traditions with *The Republic*, followed be subsequent positive utopias, the other being the social contract tradition, inspired by Hobbes, Locke, Rousseau, and Kant.

For him, even Plato's *Republic* as a whole consists of a series of small TEs, integrated into a large unity of a scenario of the ideal city; almost like the project of building a state in the armchair. The Ring of Gyges is, for Miščević, "a broadly ethical TE" (Miščević 2013, 198), in which, at the end, the nature of justice is answered by "observing in the logos how a city comes into being" (Miščević 2013, 198).

Possible Criticism: Wilkes and Sorensen

At least two criticisms have been offered to show why the Ring of Gyges is not a successful example of a TE; both are problematized by Gendler (1996). Indeed, stressing possible failures does not imply they were not "conceptually" developed as a TE; but at least they are not a good example of it. The first one comes from Kathleen Wilkes (1988), suggesting that the case is simply underdescribed, the second from Sorensen, claiming that Plato's case is incoherent. In order to demonstrate that the Ring of Gyges is a fruitful example of a TE, we should somehow reject these objections:

> The thought experiment will work (that is, provide informative data about whether fear of punishment is the only thing that leads to just behaviour) only if we already (in some sense) know how someone would act under such circumstances—if we can work it out from information we already have by recombining it in a novel but determinate way. (Gendler 1996, 233)

In the case of Wilkes, there is 'a substantial difference' in the case of the owner of the ring. She believes he must not be only invisible, but also intangible:

> If he is not intangible, he might, by mistake, bump up against someone, and get arrested by a policeman, or get his hand slammed in a till drawer. Thus a potential criminal may yet have self-interested reasons for staying within the bounds of morality /.../ If you are both invisible and intangible /.../ could you hold a gun, or a caseful of banknotes? /.../ Would others know that one owned such a ring? If so, then there might be extra reason for remaining moral: viz. that unsolved crimes might otherwise be ascribed to you /.../ The point is that /.../ the background is inadequately described, and the results therefore inconclusive. (Wilkes 1988, 11)

The criticism is possible because Plato's mythical story is underdescribed: we lack sufficient grounds for predicting the behaviour of the ring-owner because some of the crucial features of the situation remain unspecified. Thus we cannot predict what an invisible, but tangible agent would do—and therefore we are unable to make a judgment about this case. Unless we introduce two different TEs, we fail to specify whether the ring provides intangibility as well as invisibility. The question is whether such an objection is relevant to the force of the example. Gendler thinks it is not because "underdescription is a fatal flaw only if it is unresolvable" (Gendler 1996, 237).

There is another objection: we do not quite know what the world would be like for someone who is invisible and intangible. She asks, as quoted

above: could prison walls hold you? And if they could not, could you hold a gun, or a caseful of banknotes? Wilkes tries to imply that we simply do not know how one is to act under such conditions. Finally, we cannot really imagine the scenario in a persuasive way. If this fails, then the TE was truly not successful.

Gendler (1996) also refutes this case as an irrelevant remark "on a minor biological mistake in the construction of the story" (Gendler 1996, 240). The problem at hand is in the possibility of imagining what it would be like to act unobserved—we do not know even what it would be like to observe (or to see) unobserved. What seems plausible in both cases is that we cannot know in advance what role impunity plays in our decision to be moral. TEs are meant to be a tool for finding out the answer. What happens is that TEs should serve this purpose, illustrating the imaginary scenario where impunity is tracked almost perfectly, but in a way that is also precise enough. Sorensen concludes that the story is unimaginable, because it is incoherent, and for Wilkes it is as underdescribed—and for the same reason also unimaginable.

Of course, the defender of the Ring of Gyges conceived as an example of a TE can still insist that we do not need to create a perfect case for claiming something is a reliable guide to our intuitions and hypothetical reasoning. Incompleteness does not disqualify the thesis—the fatal objection must show that Plato's case lacks some central features of the very concept of a TE. Sorensen (2016) also posits the question of Plato's possible views on TEs in terms of his recollection theory as a model, claiming that it is not possible to assign any justificatory role to imagery. The produced Platonic images are instead triggers, the imagination is more a kind of memory aid—in this sense Plato counts among sceptics about TEs, he concludes, since a TE purports to justify, not merely trigger answers. Besides, Plato believes that pictures of some kind are imitations of copies, and consequently do not have any bigger educational value than things from the phenomenal world. Two problems remain unsolved, following Sorensen: in the case that the mental image in a TE is a picture, we only get a quantitative advantage over perception; there will be a richer array of triggers than perception would yield, but nothing else. There is no justification added. Another problem concerns the extension of a mental image in TEs from the field of art; since Plato famously complained about their value, the same could be true in the case of TEs. In my view, both Wilkes and Sorensen rightly demonstrated in what sense the Ring of Gyges manifests at least one feature of unimaginability. Maybe this is not a sufficient or relevant cause for rejecting the purpose of TE, but it shows at

least some vulnerability of taking Plato's story as a paradigmatic example of it.

Let us move to the interpretative objection of reading the myth of Gyges. According to most interpreters or scholars, there is supposedly a big difference between a classical and a non-classical reading. Brown somehow directly admits that his definition of a TE is rather empirical: "We know them when we see them, and that's enough to make talking about them possible" (Brown 2004, 25). Of course, the recognition of the Ring of Gyges as a TE will depend on a positive definition. The dilemma is what to do with features explicitly listed: (a) they are carried out in the mind; (b) they involve something akin to experience; (c) we typically "see" something happening in a thought experiment; (d) there is more than mere observation in them (such as calculating, application of a theory, guesswork, etc.); (e) sometimes, they contain idealizations (see Brown 2004, 25). He does not mention any relevant background conditions that need to be present before we can draw any conclusion from the imagined phenomenon.

For Miščević (2012), the Gyges TE is "aimed at discovering psychological facts". Robinson (1953, 221-222), like Sorensen, also emphasized a long time ago that Plato's use of images is condemned by his own views on images and imitation. He rejected the hypothetical method of the Ring of Gyges and saw it as inconsistent with the fact that the methods that are mostly employed in the middle dialogues are analogy and imagery. Also, demand for absolute, certain knowledge seems to be inconsistent with the recommendation of a hypothetical method that can be only approximate—a deliberate attempt is made to overcome this in the allegory of the divided line. From internal incoherencies of his usage of images, it would follow that any construction of a TE is not really possible.

Gyges's Ring could be a good example of fiction and the role of fiction in Plato. As Laird puts it, "/…/ for Plato, in *The Republic* at least, philosophical argument is principally applied to deal with cases its speakers raise that are hypothetical, and indeed fictional" (Laird 2001, 23). The passage reminds us that Socrates "has argued throughout, right up to this point, that even someone in the unlikely position of Gyges should behave justly; from now on he is going to apply the more conventional wisdom of the 'realistic response' to show that ordinary people will benefit by acting justly."

Concluding Remarks

Rescher believes that TE always rests on suppositions, drawing instruction from a process of hypothetical reasoning that proceeds by eliciting the

consequences of some projected supposition which, for aught that one actually knows to the contrary, may well be false, "/…/ such a process consists in reasoning from a supposition that is not accepted as true, and perhaps is even known to be false, but is assumed provisionally in the interests of making a point or answering a question" (Rescher 2005b, 3). In Presocratics and Plato we can find many refutatory uses of something that seems to be somewhat similar to TE. Some other uses of *modus tollens* arguments would sometimes involve their characteristic appeal to imaginary situations in order to refute a theory. Philosophers frequently appeal to imaginary situations in order to suggest their conclusions and use pre-philosophical intuitions about imaginary scenarios.

Examples from Presocratics demonstrate to us how easily we connect TE simultaneously with hypothetical or analogical reasoning on the one hand with using imagination in philosophy on the other. Are TEs sometimes only paradigmatic examples of philosophical fiction and nothing else? As a philosophical activity, fiction itself has always been a form of philosophy. The question is whether the fiction is only supplementary, or does it have a more fundamental part to play in constituting and developing philosophical arguments. To what degree can philosophical reasoning and fiction be quite generally interdependent? Much depends on what is to be understood by philosophical discourse, and whether fictions and other forms of literary creations are regarded as a natural element of it. In any case, TEs have to be much more than simple exercises in fictional activity. The Ring of Gyges, for example, needs reconsideration and a possible reconstruction of the myth in terms of a philosophical metaphor where a hypothesis, made up for the sake of argument, is changed into an act of something imaginable and a philosophical speculation is provided with some kind of fictional situation.

I was interested in determining whether the story is closer to a classic example of a TE, depicting an imaginary situation or event for the purpose of testing our common sense beliefs, or if it is better understood as Plato's account about fictional invention within philosophical arguments or perhaps his standard uses of analogy and imagery, i.e. dominant methods employed in his middle dialogues. *The Republic* is a political TE, claims Miščević (2013); his attempt to construct Plato's *magnum opus* is far more complex, taking it as an example of a political TE and a macro-TE, consisted of micro-TEs being arranged ideally into a coherent whole; at the end the imagined polis, Kallipolis, is a macro-TE consisting of a series of micro-TEs. He quotes Aristotle's criticism of Plato in the historical beginnings of the debate: "It is proper, no doubt to assume ideal conditions, but not to go beyond all bounds of possibility" (Aristotle *Politics*, II: 1265a 18). The same issue is at stake in the case of the Ring of Gyges: was Plato's idea

about picturing the consequences of invisibility really meant to be something else and more than just a mythically inspired exercise in our imagination? Maybe the interpretation about the ethical TE went beyond the bounds of his possibility, hypothesizing in a way not exactly close to his philosophical or conceptual scheme (Vezjak 2017). On the other hand, Becker (2018), for example, analyses Plato's many literary devices and styles in order to ask how Plato's fictional creations compare to TEs and how we are supposed to learn from them. Becker also identifies three main fictional styles in Plato's *Republic*—myth, simile and dialogue, relating these to the myth of Gyges, the myth of Er and the allegory of the cave; the myth told to the citizens of the model city, the construction of the model city, and *The Republic* itself as a dialogue.

A bigger problem arises with scientific ambition and stricter demand for reliability. TEs as a classic tool, enabling us to explore often impossible situations and predict their implications and outcomes, is somewhat different than a more philosophical perspective, where the conception of a TE is inherently tied in with the connection between conceivability and possibility, as suggested by the conceivability criterion of possibility by David Hume. Yet changing Aristotle's instructions about the conditions that are necessary in the case of the Platonic state that is to be constituted in the ideally best manner into methodological context, his suggestion in the case of TE could be formulated as follows: "We must therefore posit as granted in advance a number of as it were ideal conditions, although none of these must be actually impossible" (Aristotle *Politics*, VII, 1325 a 38).

We have quite a few problems when comparing TEs in natural science and in philosophy. TEs introduced by Galileo with a specific sort of evidential significance linked to the Western notion of scientific experiment are clearly different than Plato's usage of imagination in a way that is not the same as Galileo's. We can see that, in the pregnant sense of the term, it is somewhat inaccurate to speak of TEs, at least in some cases of Presocratic philosophers and even in the case of the Ring of Gyges. If we soften our interpretation of the notion and identify it by making general philosophical assumptions based on hypothetical scenarios and/or imaginative exercises, then the initial dilemma would probably quickly become redundant.

Bibliography

Aristotle. 1944. *Politics. Aristotle in 23 Volumes*. Translated by H. Rackham. Cambridge, Massachusetts: Harvard University Press; London: William Heinemann Ltd.

Bealer, G. 1998. "Intuition and the Autonomy of Philosophy". In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, edited by Depaul, M.R. and Ramsey W., 201-240. Lanham: Rowman and Littlefield.

Becker, A. 2018. "Thought Experiments in Plato". In *The Routledge Companion to Thought Experiments*, edited by Stuart, M., Fehige Y. and J. R. Brown, 44-57. New York: Routledge.

Brown, J. R. 1991. "Thought Experiments: A Platonic Account". In *Thought Experiments in Science and Philosophy*, edited by Horowitz T. and Gerald J. Massey, 119-128. Rowman and Littlefield: Savage, MD.

Brown, J. R. 2004. "Why Thought Experiments Transcend Empiricism?". In *Contemporary Debates in Philosophy of Science*, edited by C. Hitchcock, 23-43. Oxford: Blackwell.

Brown, J. R., Fehige, Y. 2016. "Thought Experiments". In *The Stanford Encyclopaedia of Philosophy*, edited by E. N. Zalta. Spring 2016. URL= <https://plato.stanford.edu/archives/spr2016/entries/thought-experiment/>.

Burnet, John. 1920. Early Greek Philosophy (3rd ed.) London: A & C Black.

Dennett, D. 1984. Elbow Room: The Varieties of Free Will Worth Wanting. Oxford: Clarendon Press.

Gendler, T. S. 1996. *Imaginary Exceptions: On the Powers and Limits of Thought Experiment*. Dissertation. Harvard, New York: Garland Press.

Häggqvist, S. 2009. "A Model for Thought Experiments". *Canadian Journal of Philosophy*. Vol 39, No. 1: 55-76.

Ierodiakonou, K. 2011. "Remarks on the History of an Ancient Thought Experiment". In *Thought Experiments in Methodological and Historical Contexts*, edited by K. Ierodiakonou and S. Roux. 37-50. Leiden: Brill.

Ieordiakonou, K. 2018. "The triple life of ancient thought experiments". In *The Routledge Companion to Thought Experiments*, edited by M. Stuart, Y. Fehige and J. R. Brown, 31-44. New York: Routledge.

Irvine, A. D. 1991. "On the Nature of Thought Experiments in Scientific Reasoning". In *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and Gerald J. Massey, 149-165. Rowman and Littlefield: Savage, MD.

Kirk, G. S. and Raven, J. E. 1957. *The Presocratic Philosophers*. Cambridge: Cambridge University Press.

Laird, A. 2001. "Ringing the Changes on Gyges: Philosophy and the Formation of Fiction in Plato's Republic". *The Journal of Hellenic Studies*, Vol. 121: 12-29.

Miščević, N. 2012. "Plato's Republic as a Political Thought Experiment". *Croatian Journal of Philosophy*, Vol. 12, No. 2: 153-165.

Miščević, N. 2013. "Political Thought Experiments from Plato to Rawls". In *Thought Experiments in Philosophy, Science, and the Arts*, edited by M. Frappier, L. Meynell, and J. R. Brown, 191-206. New York: Routledge.

Plato, *The Republic*.

—. 1968. Translated by Alan Bloom. Basic books.

—. 1969. Translated by Paul Shorey. Harvard University Press.

—. 1992. Translated by G. M, A. Grube, revised by C D C Reeve. Hackett.

—. 1991. Translated by B. Jowett. B. Vintage Classics edition.

Rescher, N. 1991. "Thought Experimentation in Pre-Socratic Philosophy". In *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and G. J. Massey, 31-42. MD: Rowman and Littlefield Publishers.

Rescher, N. 2005a. *Cosmos and Logos. Studies in Greek Philosophy*. Berlin: Walter de Gruyter.

Rescher, N. 2005b. *What if? Thought Experimentation in Philosophy*. New York: Routledge.

Robinson, R. 1953. *Plato's Earlier Dialectic*. Oxford: Clarendon Press.

Sorensen, R. 2016. "Thought Experiment and Imagination". In *The Routledge Handbook of Philosophy of Imagination*, edited by A. Kind, 420-436. London and New York: Routledge.

Tuominen, M. 2009. "Problems with 'Consciousness'". In *Approaches to Language and Cognition*, edited by H. Tissari, H. Research Unit for Variation, Contacts and Change in English (VARIENG), Vol. 3 (Studies in Variation, Contacts and Change in English; vol. 3). Helsinki: University of Helsinki.

Vezjak, B. 2017. "The Ring of Gyges and the Philosophical Imagination". In *Thought Experiments between Nature and Society: A Festschrift for Nenad Miščević*, edited by B. Borstner and S. Gartner, 410-424. Newcastle upon Tyne: Cambridge Scholars Publishing.

Wilkes, K. V. 1988. *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.

CHAPTER FOUR

CHRYSIPPUS, CYLINDER, CAUSATION AND COMPATIBILISM[1]

DANILO ŠUSTER

Part One

The compatibility of *fate* with human responsibility was one of the dominant themes in Stoicism, the debate is still with us and it is not likely to go away (according to contemporary "mysterianism", the solution might even be beyond our understanding). Terminology has changed, "fate" is out of vogue in serious philosophy, causal or even general determinism is more in accordance with modern usage. But modern interpreters suggest that Chrysippus (c. 280–207 BCE) already maintained a universal causal determinism in more or less modern sense. Frede, in her survey article on Stoic determinism, states that "No distinction between fatalism and causal determinism will be made here, in view of the fact that the Stoics etymologically derived heimarmenê, their standard term designating fate, from eirô = 'to string together'" (Frede 2003, 184; fn. 10). I claim no expertise on Stoicism, so I will defer to experts here and elsewhere. I do not have much to say about the notorious "providential" dimension of Stoic determinism either. Everything that happens in the world is, according to Stoicism, a coordinated network of causes, effects, events, and objects, "it all occurs in accordance with the plan of Zeus, and it is all bound to occur, by the bonds of Necessity" (see Brennan 2005, 235).

I will work with the most general causal interpretation of the fate principle, "everything happens in accordance with fate" (Bobzien 1998, 10). On this view, every state and event, including our actions and their

[1] A revised and expanded version of the paper presented at "Philosophical Imagination, Thought Experiments and Arguments in Antiquity", Maribor, October 9th-10th, 2018. The author acknowledges the financial support from the Slovenian Research Agency No. P6-0144.

psychology, is necessitated by prior causes (Salles 2005, xiii). Given that there are no motions without causes, the fate principle dictates that "if in identical circumstances someone will act differently on different occasions, an uncaused motion is introduced", which is metaphysically unacceptable (O'Keefe 2016, 242). And modern determinists would agree with Stoics that, when a person acts, if the internal as well as the external conditions of the person are the same, the person will invariably act in the same way. If the outcome is different in seemingly identical circumstances, there must be some hidden difference either in the external conditions or in the person's inner makeup (Frede 2003, 193).

The familiar question is then raised: if everything happens by fate, then our actions are also "bound to occur, by the bonds of Necessity" and not up to us so, "neither praise nor blame, nor honours, nor punishments are just" (Cicero *Fat*. 40; Bobzien 1998, 245). Chrysippus' reply, as reported by Cicero, is the central topic of my discussion and these passages contain the core of Chrysippus' compatibilism according to Bobzien. Chrysippus uses the illustration of the cylinder and cone and I will try to situate this analogy in the wider context of thought experiments. Historically, the analogy was not interpreted as a particularly convincing way to defend compatibility of responsibility and causal determinism. Even some contemporary compatibilists (see Dennett 1984, 2) would very likely say that, together with other imagery of the allegedly Stoic origin (a person is "dragged" by her destiny like dog tied to a cart), it is hardly more appealing than the dire prospect they are supposed to keep at bay (crude fatalism and its denial of moral responsibility). Yet I think that the analogy is still a valuable compatibilistic tool, it invites us to perform a certain "mental distillation" in which we *separate*, in imagination, the causal factors that are relevant for moral responsibility from those which are not. I will propose, somehow anachronistically, but in line both with contemporary compatibilism and Chrysippian views on causal structure of agency, that the actions of an agent are "fated" from the theoretical point of view, but they are up to her from the practical point of view.

Part Two

The causal drama of morality takes place in our minds, so let me first very briefly summarize the Stoictheory of action (Salles 2005, 34-36). The sequence leading to an action begins with an external impressor which causes the formation of the corresponding impression. The impression must subsequently be given (the internal) assent of the agent, the act of assent then constitutes (or causes) an impulse for acting in a certain way. The

practical impulse finally leads to the action itself. Crucially (Bobzien 1998, 240)—the power of either confirming or disconfirming the impression, i.e. giving assent or withholding such confirmation, is situated between the impression and our final reaction. Sometimes the internal structure is described as an assent to an impression that leads to an impulse to act (O'Keefe 2016, 240), but according to other sources every impulse actually *is* an assent (Sorabji 1980, 80). These finer distinctions will not be important, in accordance with the tradition I will discuss the sequence: an (externally induced) impression leads to assent (impulse) and assent leads to action. Any attribution of moral responsibility for our actions presupposes assent and if all of the "motions" of our mind are fated, how could we be held responsible? To give an example, if the sight of beauty by necessity provokes love in an uncontrolled man, how could his actions be up to him (see Frede 2003, 192)?

As reported by Gellius, "He [i.e. Chrysippus] then uses an illustration of this point which is quite suitable and witty" (Bobzien 1998, 259). The example of the cylinder and the cone ("C&C" for short) is usually described as an *analogy* (O'Keefe 2016, 240; Bobzien 1998, 258; Salles 2005, 44; Frede 2003, 194), and less often as a *simile* (Eliasson 2008, 88). Let me quote Cicero,

> For when it is said that assents happen by means of preceding causes, Chrysippus believes that he can easily explain how this works. For, even though an assent cannot occur unless set in motion by an impression, none the less, since the assent has this impression as proximate cause and not as principal cause, it has the reason, as Chrysippus holds, which I stated earlier: it is not the case that assent can happen without being prompted by some force from outside—for it is necessary that an assent be set in motion by an impression but <in order to make this clear> Chrysippus returns to his cylinder and cone, which cannot start moving without being pushed. However, when this has happened, he believes that from then on the cylinder rolls and the cone spins by their own nature.
>
> Thus, he says, just as the person who shoved the cylinder gave it the beginning of its motion, but did not give it its rollability, so likewise, an impression, when encountered, will imprint and so to speak stamp its form on the mind, but assent [to it] will be in our power; and, just as was said in the case of the cylinder, being pushed from outside, for the rest it will move by its own power and nature. (*Fat*. 41-3; Bobzien 1998, 258-59)

The analogy suggests that the movement of the cylinder is a result of two types of causal factors—one is our pushing the cylinder to make it move (the initiating or "auxiliary and proximate cause", often described as the *external* cause). And the other is the cylinder's being round, its "nature"

(usually described as "the perfect and principal" cause), the "internal" causal factor. In the same way our assents, and consequently our actions, are the product of two causal factors, external stimuli from the environmental circumstances ("impressions") and internal reactions determined by the state of our minds, our characters. The impression (not "up to us") is the externally induced antecedent causal factor of the act of assent. Different human beings might be presented with the same *stimulae* (the external antecedent cause, e. g. "the sight of beauty") but they will react differently, according to their mental dispositions. So, the nature of the objects at which the effect takes place (geometrical bodies, human beings) must be causally responsible for the differences in effect. The finer mechanics of the causal web is not so clear, however, Cicero (*Fat*. 41) reports

> Among the causes some are complete and principal,2 others auxiliary and proximate. For this reason when we say that everything happens by fate through antecedent causes, we do not want this to be understood as if it were through complete and principal causes, but through auxiliary and proximate ones. (quoted from Frede 2003, 187)

According to the standard interpretation, a distinction between the antecedent or *external* and the principal or *inner* cause explains how human beings are part of the universal causal web in a way that leaves room for personal responsibility. The C&C analogy is supposed to limit the role of the external *antecedent* cause in the mechanism that leads to the assent and subsequently to action. What is responsible for the rolling of the cylinder and spinning of the cone is their nature manifested in their disposition to move in a particular way. Although antecedent causes initiate every process, they are only necessary and never sufficient, they are not involved in the ensuing activity itself. The characterizations of causes as antecedent/external/proximate/auxiliary go together and so do principal/perfect (complete)/internal/the main causal factor. But several questions immediately emerge: are complete (perfect) and principal (primary) always paired or not? How about auxiliary and proximate? Are there any further subdivisions? Can an antecedent cause also be the principal cause (and not just auxiliary)? And how do necessity and fate qualify the nature of links in the causal web?

The relevant passages from *De Fato* are not easy to understand. Bobzien even remarks "/…/ structurally, Cicero's *On Fate* is rather a mess /…/ (drawing from several sources, perhaps in some haste and without final

[2] "Perfectae et principales" in Latin – "perfect and principal" according to Bobzien (1998, 256); "perfect and primary" according to Salles (2005, 42).

editing)" (Bobzien 1998, 332). According to the dominant interpretation (Frede, 2003), the Stoic compatibilism consists in making the internal causes the *principal* causes in determining the action. Two determining factors *cooperate* in bringing about the effect (the act of assent). The *internal* determining factor is a perfect and principal cause which produces the effect in cooperation with the *external* auxiliary and proximate cause. The opponent is depicted as saying that fate (operating through "uncontrollable" impressions) is a necessitating and constraining force external to the agent. Chrysippus agrees that "everything comes about by fate according to antecedent causes." But antecedent causes are auxiliary and proximate only and as such they do not render their effects *necessary* (at least in the cases at issue). The actions depend on the agent because that which determines the quality of the relevant effect (assent, action) is the inner nature of the agent's mind. Assent can be withheld and the C&C analogy suggests that different people will react differently to comparable externally induced stimuli (the cylinder and cone can be seen as representing good and bad moral characters). In the determination of human action fate works *through* human beings—the dispositions of our minds are subject to fate in a particular way, namely "in such a way as to accord with their characteristic quality" (Bobzien 1998, 251). The subsequent action is not necessary because it is not externally forced, but it remains "fated"—like everything else it has antecedent causes in its causal history.

Part Three

Historically, the passages from *De fato* (41-5) present the most discussed text on Stoic "compatibilism." The prevailing view was that the C&C analogy is not a very convincing way to defend compatibility of responsibility and causal determinism. The reluctance to see ourselves compared to rolling cylinders and spinning tops is perhaps understandable. True, there are also some assenting voices regarding "the cylinder of Chrysippus". Leibniz, for example, remarks, "He is right in saying that vice springs from the original constitution of some minds" (*Theodicy* §335; quoted by Forman 2016, 232). But Leibniz and the idea that I am free whenever the cause of my action is within me is famously ridiculed by Kant as a "wretched subterfuge",

> /…/ and if the freedom of our will were none other than the latter [kind] (say, psychological and comparative freedom, not simultaneously transcendental, i.e., absolute, freedom), then it would basically be no better than the freedom of a turnspit, which, once it has been wound up, also performs its motions on its own. (Kant 2002, 123)

This type of criticism, apparently based on the mechanistic nature of the analogy, (probably) began with Alexander of Aphrodisias—to him the very fact that the model applies to non-rational and rational beings alike represents its major flaw (Frede 2003, 193). But I think that this diagnosis is based on a misapprehension of the role of the C&C analogy. Analogies are often the basic ingredient of a thought experiment (TE), so how should we understand the imaginary scenario used by Chrysippus in relation to TE? Let me start with a very minimalistic characterization of a TE, "to perform a thought experiment is to reason about an imaginary scenario with the aim of confirming or disconfirming some hypothesis or theory" (Gendler 2004, 1155). We begin with a question or a hypothesis to be tested. The challenge faced by Chrysippus is easy to formulate: how is it possible for an action to be fated (by way of antecedent causes) and still up to us (in a morally relevant sense)?

Well, in his reply "Chrysippus returns to his cylinder and cone" (*Fat*. 42.3; Bobzien 1998, 259). Is the analogy exploited as an *argument* aimed at confirming or disconfirming some hypothesis or theory? Perhaps: "Case #1 (the movement of the cylinder) is like Case #2 (our assent); in Case #1 the movement is explained by the nature of the body rather than by the initial push, so in Case #2 our assent is explained by the disposition of our own mind rather than by external causes". The argument is rather short, to be sure, but argumentative analogies can be very short. Govier quotes from a letter to the editor about gun control: "guns are no more responsible for criminal deaths than forks are responsible for obesity" (Govier 2016, 1). Forks (in the Western world, at least) are part of the causal background leading to but not responsible for obesity. *So* guns are not responsible for criminal deaths either. The problem with C&C is not its shortness but something else: if you are not *independently* convinced that the cylinder moves because of its own nature (and not because of the initial push) then the analogy offers little to persuade you. This is a *figurative* analogy (Waller 2001, 200) and figurative analogies illustrate and (sometimes) elucidate, but they do not offer any reasons or argue for a conclusion.

There are at least two types of analogies, argumentative and figurative, but there are further subdivisions. Let me bracket argumentative analogies (deductive and inductive) and just consider the figurative variety. A figurative analogy typically uses more familiar images (the source case) to help us understand something that is complex, confusing, or unfamiliar (the target case). But not all figurative analogies are alike, some of them are *illustrative* only, they offer compelling images, while others are also *explanatory*, the source displays some important structural features of the target. The distinctions are often blurred, in a certain sense all figurative

analogies are illustrative, but some are just more or less elaborate metaphors or comparisons that make certain characteristics of the target more vivid in accordance with the working definition of a *metaphor*: "seeing, experiencing, or talking about something in terms of something else" (Ritchie 2013, 8). In *Averroes' Search*, for instance, Borges (1999) discusses how Zuhair (Arabian poet) compared destiny to a blind camel. The story tells us that "in the course of his eighty years of pain and glory many is the time he has seen destiny trample men, like an old blind camel" and then gives a more elaborate explanation, "/…/ every man has surely felt at some moment in his life that destiny is powerful yet clumsy, innocent yet inhuman. It was in order to record that feeling, which may be fleeting or constant but which no man may escape experiencing, that Zuhair's line was written" (Borges 1999, 240).

Let us compare this metaphor about destiny to another famous (and depressing) Stoic simile—when a dog is tied to a cart, if it wants to follow, it is both pulled and follows and if it does not want to follow, it will be in any event necessitated. And the same holds for human beings—even if they don't want to, they will be compelled to follow what is destined. The simile is open to several interpretations but one can easily understand that it historically provoked the resistance against Stoic doctrines—with advocates like that, why do *compatibilists* need any enemies at all? Bobzien actually thinks that there is no reason to assume that the dog-cart analogy is Chrysippean at all (Bobzien 1998, 357). But I am only interested in the structure of the comparison: it seems to me that the dog and the cart image functions like Borges' "blind camel" metaphor: an illustration that draws our attention at aspects of the phenomenon that we try to elucidate by means of comparison and perhaps records "a certain feeling".

Not so with the C&C analogy—very modest from an artistic point, yet it still succeeds in displaying the relevant structural properties of the target case. Let me introduce another famous case. When trying to understand the mysterious world of quantum phenomena and the nature of light (the wave-particle duality), one sometimes encounters, as an illustration, a *cylinder* that we cannot perceive globally "as it is" but only in one of its aspects. A cylinder is circular as seen from one angle (base), and rectangular from another angle (looked from the side). The *quantum* object, by analogy, is one that we can perceive *via* experiments that show only one of its aspects, but in reality it is something more complex, more than just a circle plus a rectangle. The authors of a typical textbook on modern physics remark, rather dryly, "This metaphor is very interesting, but it does not explain anything scientifically and logically" (Yadav et al. 2014, 52). Well, in the area of moral responsibility, we do not really expect scientific equations and

we are satisfied with explanations displaying the relevant causal structure, and this is precisely the merit of the C&C analogy.

The Stoic distinctions in the area of causality are perhaps confusing but it seems clear that the causal background of the act of assent consists of (at least) two types of causal factors. In terms of modern discussions, we often want to distinguish causes from mere conditions. Mackie introduced a useful notion of a causal field (Mackie 1980, 34). Causal statements are commonly made against a certain background and a causal statement is the answer to a causal question of the type: "What caused this X?" This question can be expanded into: "What made the difference between those times, or those cases, within a certain range, in which X did not occur and the case when X did occur?" But how to differentiate between real, active causes and mere background conditions in the causal field? David Lewis is famous for saying:

> We sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others. Or we single out a few as the 'causes', calling the rest mere 'causal factors' or 'causal conditions'. Or we speak of the 'decisive' or 'real' or 'principal' cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. (Lewis 1986, 162)

Apparently we choose the most *salient* features as the real causes and the question arises as to whether there is any objective difference between causes and conditions. Suppose we agree that the choice is contextual, dictated by our explanatory interests. It does not follow that the selection is made in a totally capricious manner. Mackie gives an example of a fire "produced" by a combination of a short circuit and the presence of oxygen. Most speakers will distinguish between the short circuit as "the cause", and the presence of oxygen as a mere "background condition". Such a selection is also integral to our moral and legal practices. We know whom to blame if the old electric installation was preventively checked a week before. What made the difference in the case of fire was, say, a careless examination of the installation and not the poor condition of the wires by itself. In their classic work on causation (in law), Hart and Honore stipulate that *abnormal* factors are treated as causes—a digression within a presupposed normal causal field as Mackie would say (oxygen is *usually* something to be expected and a short circuit is not). They also maintain that free deliberate actions are always (proper) causes, while normal conditions and non-agential factors are conditions only (Hart and Honore 1959, 31).

In our case, Chrysippus was faced with the challenge—if everything happens by fate and antecedent causes, then we would have to say that our

choice is caused by something not in our power. As a reply he offers an explanatory analogy exactly in line with Hart and Honore: the external antecedent causes form only a part of the causal field, the inner, agential factor (our mental dispositions) is the *explanatory* superior factor which leads to the effect. The analogy is persuasive in the sense of making the characteristics of that aspect of the target object or phenomenon (mental processes that take place in the mind) that we try to illustrate more vivid by means of the comparison (a succession of physical events on the level of perceptible everyday objects). TEs are often based on arguments from analogy, but they also exploit the explanatory features of analogies. One of the main features is exactly the one exhibited by the C&C analogy: "mental" filtration of ingredients of the causal field into distinct elements and separating in imagination of the morally relevant aspects (see Miščević 2012, 202). This "separating" in "experiment with our mind" of elements that normally go together is the reason that the simple C&C analogy functions as more than a plain metaphor.

I think that the mechanical, "turnspit" interpretation of the C&C analogy misses the point of the *explanatory* analogy. Bobzien (1998, 260) is right to observe that not every detail of the *explanandum* level has a parallel on the *explanans* level of the analogy or vice versa—the non-rational mechanic nature of geometrical bodies is just irrelevant. Moreover, Chrysippus does not offer an *argumentative* analogy. If a thought experiment is an imaginary scenario explored for the purpose of acquiring knowledge (Rescher 2005, 61–72) then the C&C analogy is clearly not a TE. It is not as if we *mentally* roll a cylinder and then draw a certain conclusion about the target case (a child can easily roll a tin can on an inclined surface). One would have to give *independent* reasons for the special treatment of the internal, agential factors. Hart and Honore, for instance, just take the existence of free deliberate actions for granted when they claim that free agents should always be treated as proper causes in the causal background. Explanatory analogies are persuasive in the sense of making the characteristics of that aspect of the target object or phenomenon that we try to illustrate more vivid by means of the comparison. The analogy separates the main causal factor from the total causal mixture (causal field), so to speak. We are led to "see" the "roundness" of the cylinder as the primary *explanation* of its rolling and in the same way the assent (action) is explained by the disposition of our own mind rather than by external causes. Thus the foolish, "go astray through their own impulse and are harmed by their own purpose and determination" (Chrysippus according to Gellius, as quoted by Forman 2016, 232).

Part Four

What is the modern relevance of Chrysippian compatibilism? Bobzien remarks that the analysis of the relevant texts "has yielded little that bears any resemblance to *modern* arguments for the compatibility of causal determinism with freedom" (Bobzien 1998, 276). No Greek word for freedom is used in the passages. According to influential modern interpretation, to say of someone that she "has free will" or "has a choice" with respect to a certain action at a given time, is to grant her an ability to act in that way and an ability to do otherwise (van Inwagen 1983, 162). But "we have no reason to assume that Chrysippus, or his opponents, were involved in a debate about the compatibility of freedom to do otherwise and causal determinism, nor that they based moral appraisal directly on the idea that the agent could have done otherwise" (Bobzien 1998, 279). The diagnosis is surprising, just consider the famous passage:

> If everything happens by fate, everything happens by way of an antecedent cause. And, if impulse, so too those items which follow impulse, hence also assents. And, if the cause of impulse does not lie with us, neither does impulse depend on us. But if this is so, those items, too, that are the effect of impulse do not lie with us; therefore neither assents nor actions depend on us. From which it follows that neither praise nor blame, nor honours, nor punishments are just. (*Fat.* 40; Bobzien 1998, 245)

The causal sequence: "impulse, thereafter assent" is confusing since we earlier treated the act of assent as constituting or causing an impulse for acting in a certain way. Bobzien thinks that we actually have a sequence: (i) external object; (ii) impulsive impression; (iii) assent to the impulsive impression; (iv) action (Bobzien 1998, 247). I must once again defer to the experts, but it seems clear that the passage presents an argument for incompatibilism which is surprisingly similar to the main contemporary argument for the compatibility of free will and determinism:

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (van Inwagen 1983, 56)

Antecedent causes are now explicated more generally in terms of laws of nature and the state of the world at the distant past. The central notion in the antique debates was "that which depends on us", "that which is in our power", and they would not explicate "that which is up to us" in terms of

our ability to act otherwise. Still, the raw structure of the two arguments is basically the same:

X is not up to S.
S's action is a causal consequence of X.
So,
S's action is not up to S.

Next, consider the *main* compatibilist response to this argument. Distant past and laws of nature are unavoidable (not up to us) because they are independent of our beliefs and desires. There is nothing we can do to change them, they transcend our causal abilities where the latter are defined as, say:

> S is *causally* able at t to bring about *p* iff there is a course of action K such that at t (i) S is able to do K, and (ii) S's doing K would make it the case that *p*. (Kapitan 2002, 134)

The premises (unavoidability of past and laws of nature) are true in *this* sense of ability, but the conclusion (unavoidability of our actions) does not *follow*. S's action is causally dependent upon S's motives, upon her internal make-up. Well, this is precisely the Chrysippian reply: an action can truly be said to be "up to the agent" if the agent's "nature" is the main causal explanation of her action, never mind the "fate" and the fact that the external antecedent causes are not up her.

The incompatibilists usually disagree with the causal definition of ability and perhaps describe it as contrived and *ad hoc* (van Inwagen 2002, 167) or claim that the validity of the argument for incompatibilism is much more obvious than any compatibilist analysis of the ability (van Inwagen 1983, 222-223). They are particularly critical of once popular conditional analyses: to say that, at the time of acting, S could have done Y and not X is just to say that, had she wanted (chosen, willed, or decided) to do Y and not X at that time, then she would have done Y. This line of debate was then more or less driven to stalemate with each side accusing the other of question-begging with respect to the relevant notion of (in)ability. No wonder that many contemporary compatibilists prefer to bypass the impasses of (in)ability to act otherwise and welcome the idea promoted by Harry Frankfurt (1969), which is that moral responsibility does not require alternative possibilities at all. The Stoic approaches to free will are then often claimed to be strikingly similar to Frankfurt's theory of moral responsibility.

According to Frankfurt, the responsibility for the action derives from the agent's decision to perform it and from that decision's being based on a previous all-things-considered practical reflection.

> Similarly, the responsibility for the action in Chrysippus derives from the agent's exercise of an impulse for it (or his assenting to the impression where the action is presented as valuable), but also, and crucially, from the impulse's being fully rational, which involves a reflection concerning the all-things-considered desirability or appropriateness of the action. (Salles 2005, 66)

The C&C analogy illustrates the fact that an agent may be responsible for her action even if the whole sequence is set in motion by external factors (the proximate, antecedent causes). Compare the often-quoted passage from Frankfurt:

> To the extent that a person identifies himself with the springs of his actions, he takes responsibility for those actions and acquires moral responsibility for them; moreover, the questions of how the actions and his identifications with their springs are caused are irrelevant to the questions of whether he performs his actions freely and is morally responsible for performing them. (Frankfurt 1988, 54)

The internal factors are the locus of responsibility and the agent's autonomy is not diminished by the fact that the inner realm is also subject to "fate" (causal history). Lack of external coercion and the fact that the agent acts on the basis of her best practical reflection explain the agent's responsibility.

This historical swing brings us back to the tradition that does not base moral appraisal on the idea that the agent could have done otherwise. This might suggest that instead of desert-based accounts of moral responsibility (roughly, an agent is morally responsible for X if and only if she is deserving of praise or blame in virtue of having freely brought about X), a Chrysippian should opt for consequence-based accounts of moral responsibility (an agent is morally responsible for X just in case praising or blaming her for X would produce good consequences, see Klampfer 2014). But I am not sure how Stoicism as a system of virtue ethics fits with modern utilitarianism, so I prefer to avoid these dilemmas. Especially since neither Chrysippus nor modern compatibilists give up the idea that an action has to be up to the agent for her to be morally responsible. But then the agent should have an option, in a certain sense, to act otherwise. Let me explain this with the help of a rather unusual (and one would say totally anti-Stoic) example. Sartre discusses the following ordinary situation:

> The alarm which rings in the morning refers to the possibility of my going to work, which is my possibility. But to apprehend the summons of the alarm as a summons is to get up. Therefore the very act of getting up is reassuring, for it eludes the question, "Is work my possibility?" Consequently it does not put me in a position to apprehend the possibility of quietism, of refusing to work, and finally the possibility of refusing the world and the possibility of death. In short, to the extent that I apprehend the meaning of the ringing, I am already up at its summons; this apprehension guarantees me against the anguished intuition that it is I who confer on the alarm clock its exigency—I and I alone. (Sartre 1992, 37-38)

It is (reasonably) clear what he means: the alarm clock goes off and I get out of bed as if I had no choice but to obey "its summons" and avoid the existential anguish of having all the options open. But what if I do not elude other possibilities and freely consider my staying in bed as an option? Is my staying in bed any less determined?

Consider the case of cosy Clara, woken up by an alarm clock but, say, "apprehending the possibility of quietism", she remains in her bed. Though she has no reason or inclination to arise, she is perfectly able to do so, according to Watson (2004, 91). We can imagine the following causal sequence: (external) comfortable impressions making Clara overwhelmed by the impulse characterized as "one's great desire to persist in the state of rest" thereby giving her reasons to stay in bed. She is unwilling to get up, she has no reason or inclination to arise, but this action is still up to her, because she stays in bed because of her own "nature". And she *has* other options: had she decided or intended to act otherwise, she would have arisen.

She would act otherwise had certain internal conditions been different. Why concentrate, with Chrysippus and modern day compatibilists, on her mental make-up and its potentials in the entire causal background of an action? Clara's *actual* inclination is just a part of the total causal field, one of the necessary conditions of her staying in bed like other *sine qua non* concerning the conditions of her limbs and muscles and central nervous system (see Watson 2004, 92). From the *theoretical* viewpoint, from the perspectives of prediction and control, all of the factors that are co-responsible for her action are equally part of the overall causal network. Making Clara very comfortable may be just as effective a means of preventing her from getting to work as chaining her to the bedpost or paralyzing her is. Compare: once set in motion, the cylinder will roll according to its own nature. Once set comfortably in her cosy environment, Clara will stay in bed, the "impressions" lead her to laziness.

But in practical contexts (deliberation, agency, responsibility), we understand what someone can do in terms of a dependency relation between the individual's motives and her behaviour. In these contexts, not all of the necessary conditions in the causal background have equal force. There is an important difference for Clara not coming to work because she was externally prevented (chained, locked or perhaps just ill) or whether she stayed at home for her own reasons. We can agree with Watson that Clara is not rendered *impotent* by her lack of interest in leaving her comfortable environment, "/…/ it would show confusion about the meaning and role of the ordinary notion of ability to say of lounging Clara that she can't get out of bed just because one of the necessary conditions of her doing so is absent—namely, her (actual) intending to get up" (Watson 2004, 93). I take it that our *ordinary* notion of ability captures the idea of what is up to me ("I can do X" corresponds to "doing X is up to me"). Clara's lack of will does not imply her lack of autonomy (in the sense relevant for moral responsibility) usually associated with one's disabilities. The unwilling Clara is perhaps unable to get up from the *theoretical* point of view, but this action is still up to her, it depends on her will.

Consider the familiar point, attributed to Spinoza, that one cannot both intend to do something and predict that one will do it at the same time (Bilgrami 2006, 251). The purpose of the *theoretical* point of view is description, prediction and explanation. Given Clara's actual state of mind, her (in)action is predictable (and in this sense necessary) given *all* of the antecedent conditions (see also Bok 1998, 62-65). The purpose of *practical* reasoning is to determine one's will, to determine what we have reason to do. Clara, when deliberating whether to get up, is not trying to predict her future from the "third person" point of view, but is determining it as an agent from the "first person" perspective on herself. She is trying to determine what to do in view of her options—those courses of action whose realization she sees as depending on her decision, choice, or intention (her "nature"). Generally speaking, when I consider various courses of action in my practical reasoning, I ask myself what would be the case if I were to choose X or not. I am not trying to predict my action, so I do not assume anything about what I will in fact choose, I am trying to make up my mind about what I have reason to do (see Watson 2004). This bracketing of actual intentions opens up the space of possibilities for the agent—Clara is not limited to the course of action determined by the antecedent conditions of her actual choice (precisely specified past and fixed laws of nature).

I would propose the theoretical/practical standpoint (Bok 1998), the viewpoint of intention or performance conditions/the viewpoint of prediction or enabling conditions (Watson 2004), or first and third person

points of view (Bilgrami 2006) as distinctly modern developments of the Stoic "fated, yet still up to us" solution of the puzzles of free will. This might look anachronistic, but I think it helps us appreciate the Stoic perspective on the causal structure of agency. I think that the general contours of the conditional analysis of ability are in line with Chrysippus and his view that the act of assent is non-necessary:

> 'I assent to this impression' could in principle be false and is not externally prevented from being false. /…/ because although the impression entered my mind, there would be no external force or hindrance that prevents the act of assent from not happening. /.../ The quality of the response depends on the nature of the agent's mind. (Bobzien 1998, 312-13)

Clara's responses to the alarm clock are not externally forced, they depend on her will and her responsibility is judged by the question: could she come to work had she chosen to do so? Was this an option?

Part Five

If an action depends on the agent, it is up to the agent and not externally forced, so not necessary but it is still fated (or necessitated but not *necessary*, for this interpretation see Bobzien 1998, 126 and 312). Still, the necessity of fate applies to every detail, in accordance with the famous remark made by Heraclitus, "Character is fate" also. This is unacceptable for libertarians. Sartre, for instance, would deny any type of necessitation or the relevance of theoretical contexts (bad faith!)—there is only the reality of agents and their unconditional ability to act otherwise. The pessimists will point out that if one's action results from a deterministic causal process that traces back to factors beyond one's control, then one is not free in the sense required for moral responsibility (Pereboom 1995).

Well, a monarchist is committed to defending a monarchy and Chrysippus, I think, is committed to (what seem to be) analytic consequences of Stoic determinism (every state and event is fated). But I do not agree that "/…/ the most one can say (with Oenomaus II 978) is that, if Democritus makes the soul of man a slave, Chrysippus makes it a half-slave" (Gould 1971, 151). I think that what grounds these pessimistic conclusions is a misunderstanding of determinism as involving threats that suggest that our rational, conscious mental activity is *bypassed* in the process of our making decisions and coming to act (see Nahmias 2011). I think that the C&C analogy has a real explanatory value in addressing these fears. The cognitive role of this simple analogy is to "isolate" the salient casual factor and help

us understand the idea of responsibility as demanding the autonomy of the agent as the cause of her own actions.

In the spirit of Stoicism, the universe and everything is just matter and motion (with or without *logos*). In the realm of agency one could say that it is all about control and control is about causation. Chrysippus pointed out that the most important node in the causal chain of our action is what we *identify* with (our intentions and our character). But we aspire for more—as agents we want to be the "prime mover unmoved". The compatibilist solution sketched above, which combines the first person point of view of a free agent and a theoretical, third person point of view of an agent as the object of causal and motivational histories, looks paradoxical. But so does the strange world of quantum phenomena (the wave-particle duality). The example of a cylinder (circular from one angle and rectangular from another) can perhaps help us understand the mysteries of quantum reality. And so, too, the old and simple image of a cylinder and a cone may still help us comprehend the causal drama of human responsibility.

Bibliography

Bilgrami, Akeel. 2006. *Self-Knowledge and Resentment*. Harvard: Harvard University Press.

Bobzien, Susanne. 1998. *Determinism and Freedom in Stoic Philosophy*. Oxford: Oxford University Press.

Bok, Hilary. 1999. *Freedom and Responsibility*. Princeton: Princeton University Press.

Borges, Jorge Luis. 1999. *Collected Fictions*, translated by Andrew Hurley. New York: Penguin Books.

Brennan, Tad. 2005. *The Stoic Life. Emotions, Duties, and Fate*. Oxford: Clarendon Press.

Cicero. 1948. "De Fato". In *De Oratore* (Vol. 2), translated by H. Rackham. Cambridge, Mass: Harvard University Press.

Cicero. 1938. *De divinatione, De fato, and Timaeus*. Edited by W. Ax. (Leipzig 1938, reprinted Stuttgart 1965). Quoted after English translations given in Bobzien (1998) as *Fat. nn.*

Dennett, Daniel. 1984. Elbow Room: The Varieties of Free Will Worth Wanting. Cambridge, Mass.: MIT Press.

Eliasson, Erik. 2008. The Notion of That Which Depends On Us in Plotinus and Its Background. Leiden: Brill.

Forman, David. 2016. "Leibniz and the Stoics: Fate, Freedom, and Providence". In *The Routledge Handbook of the Stoic Tradition*, edited by J. Sellars, 226-242. London: Routledge.

Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility". *Journal of Philosophy*, Vol. 66: 829–839.

Frankfurt, Harry G. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Frede, Dorothea. 2003. "Stoic Determinism". In *The Cambridge Companion to the Stoics*, edited by B. Inwood, 179-205. Cambridge: Cambridge University Press.

Gendler, Tamar Szabó. 2004. "Thought Experiments Rethought-and Reperceived". *Philosophy of Science*, Vol. 71, No. 5: 1152-63.

Gould, Josiah B. 1971. *The Philosophy of Chrysippus*. Leiden: Brill.

Govier, Trudy. 2016. "Outstanding Questions about Analogies". In Argumentation, Objectivity, and Bias: Proceedings of the 11 th International Conference of the Ontario Society for the Study of Argumentation (OSSA), edited by P. Bondy and L. Benacquista, 1-12. Windsor, ON: OSSA.

Hart, Herbert Lionel Adolphus and Honore, Anthony Maurice. 1959. *Causation in the Law*. Oxford: Clarendon Press.

Kant, Immanuel. 2002. *Critique of Practical Reason*. Indianapolis: Hackett Publishing Company.
Kapitan, Tomis. 2002. "A Master Argument for Incompatibilism?". In *The Oxford Handbook of Free Will*, edited by R. Kane, 127-157. Oxford: Oxford University Press.
Klampfer, Friderik. 2014. "Consequentializing Moral Responsibility". *Croatian Journal of Philosophy*, Vol. 40: 121-150.
Lewis, David. 1986. "Causation". In *Philosophical Papers 2*, 159–213.Oxford: Oxford University Press.
Mackie, John Leslie. 1980. *The Cement of the Universe*. Oxford: Oxford University Press.
Miščević, Nenad. 2012. "Political Thought Experiments from Plato to Rawls" In *Thought Experiments in Science, Philosophy, and the Arts*, edited by M. Frappier, L. Meynell and R. Brown, 191-206. New York: Routledge.
Nahmias, Eddy. 2011. "Intuitions about Free Will, Determinism, and Bypassing". In *The Oxford Handbook of Free Will* (2nd. ed.), edited by R. Kane, 555-576. Oxford: Oxford University Press.
O'Keefe, Tim. 2016. "The Stoics on Fate and Freedom". In *The Routledge Companion to Free Will*, edited by K. Timpe, M. Griffith and N. Levy, 236-246. London: Routledge.
Pereboom, Derk. 1995. "Determinism al Dente". *Noûs*, Vol. 29, No. 1: 21-45.
Rescher, Nicholas. 2005. *What If? Thought Experimentation in Philosophy*. New Brunswick, NJ: Transaction Publishers.
Ritchie, David. 2013. *Metaphor*. Cambridge: Cambridge University Press.
Salles, Ricardo. 2005. *The Stoics on Determinism and Compatibilism*. Aldershot: Ashgate.
Sartre, Jean-Paul. 1992. *Being and Nothingness: A Phenomenological Essay on Ontology*, translated by H. Barnes. New York: Washington Square Press (*L'être et le néant*, 1943).
Sorabji, Richard. 1980. Necessity, Cause and Blame: Perspectives on Aristotle's Theory. London: Duckworth.
van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
van Inwagen, Peter. 2002. "Free Will Remains a Mystery". In *The Oxford Handbook of Free Will*, edited by R. Kane, 58-177. Oxford: Oxford University Press 2002.
Waller, Bruce. 2001. "Classifying and Analyzing Analogies". *lnformal Logic*, Vol. 21, No. 3: 199-218.
Watson, Gary. 2004. "Volitional Necessities". In *Agency and Answerability: Selected Essays*, 88-122. Oxford: Clarendon Press.
Yadav, Raja Ram. et al. 2014. *Modern Physics for Scientists and Engineers*. Delhi: PHI Learning.

CHAPTER FIVE

LUCRETIUS' SPEAR THOUGHT EXPERIMENT FOR THE INFINITE UNIVERSE

MAJA MALEC

Introduction

In his beautiful philosophical poem, *De rerum natura* (*On the Nature of Things* 2001), written in the middle of the first century BCE, Lucretius mainly introduces Epicurus' natural philosophy to the Roman audience. The universe is eternal and as a whole immutable, and consists of two things, "matter and the void in which matter is located and moves in every direction" (Lucretius 2001, 1.420-421). By "matter" Lucretius actually means indivisible particles, atoms (from the Greek adjective "atomos", which means "indivisible, what cannot be divided") out of which compound bodies are built. Lucretius avoids using the Greek loan-word and instead uses a series of descriptions to emphasise the creative power of atoms; besides "matter" he also calls them "primary elements of things", "generative particles of things", "seeds of things", "ultimate particles" (1.55-61) as well as "little bodies" and "elements" (Warren 2010, 22). Atoms are solid, indivisible and thus indestructible, while compound bodies, which are also void, are destructible and sooner or later dissolve (Lucretius 2001, 1.512-519). Void is indestructible too, "everlasting". "Void" is a term by which he names an "empty space", i.e. space in which there is no matter (1.508-509), but often he simply means space, i.e. that in which bodies are located and through which they move (1.421). The existence of void—inaccessible to sense perception—is being inferred from the existence of moving bodies. As Lucretius puts it, "if there were no room and space—void, as we call it—matter could not be located anywhere, and its movement in any direction would be absolutely impossible" (1.425-429). In fact, atoms are too small to be perceived, so what we actually perceive are compound bodies, and from them we infer the existence of atoms and void.

Another important tenet of Epicurean natural philosophy is the infinity of the universe, for which Lucretius argues at the end of Book I (1.951-1117). At this point we also encounter the Spear thought experiment (1.968-983), the topic of this chapter by which Lucretius exemplifies the absurdity of the finite, bounded universe, thus proving that the universe can only be infinite. In the preceding passage (1.958-967), Lucretius argues for the infinity of the universe by exploiting the notion of limit, and according to one interpretation the thought experiment does not introduce a new argument, but rather serves as an illustration of the first one, which I will call the "Limit argument". The choice is partly dependent on a particular account of thought experiments we adopt, whether we believe that thought experiments are independent tools of investigation or just dressed up arguments, and partly on how we analyse the thought experiment. Here it is helpful to compare it with a similar thought experiment, perhaps it is even more appropriate to speak of another version of the same thought experiment. This is actually the first recorded thought experiment, attributed to the Pythagorean Archytas of Tarentum in the first half of the fourth century BCE. The aim of both is to prove the absurdity of the finite world, but while Lucretius asks us to imagine a man running towards the very edge of the universe and then throwing a spear, Archytas has a man standing at the very edge of the cosmos trying to extend his hand or his staff. Another issue that needs to be settled is whether this is actually a thought experiment. Thought experiments were properly recognised as a distinct investigative tool only in the previous century, and their characterisation actually depends on the notion of experimentation as it was established in modern natural sciences. Is it appropriate to speak of ancient thought experiments? They certainly did not call them so and did not, at least not explicitly, distinguish them from ordinary arguments.

First, I will describe the Spear thought experiment and the Limit argument. Then I will examine whether the Spear thought experiment is independent proof of the infinity of the universe or whether it is merely an illustration of the Limit argument. Finally, I will turn to the question of whether this is actually a thought experiment.

I. The Limit Argument and the Spear Thought Experiment for the Infinity of the Universe

At the end of Book One (1.951-1117), Lucretius sets out to determine "whether the void whose existence we have discovered, the place and space in which everything happens, is essentially finite, or whether it opens out to boundless breadth and abyssal depth" (Lucretius 2001, 1.955-7). He does

not let us wonder about what the right answer is for long. He immediately continues with the claim that "the universe is not bounded in any direction" (1.958), bringing forth a series of arguments for the infinity of the universe, the second actually being developed in the form of a thought experiment—the Spear thought experiment. One could see the thought experiment merely as an illustration of the first argument, i.e. the Limit argument, so it is best to describe both.

Lucretius' Limit argument, an expanded version of Epicurus' argument from the "Letter to Herodotus" (Bailey 1926, sec. 41), goes as follows:

> In fact, the universe is not bounded in any direction: otherwise it would inevitably have an extremity. Now it is plain that nothing can have an extremity, unless there is something on the farther side to bound it, so that there is seen to be a point beyond which our vision cannot trace the object. And since we must admit that there is nothing outside the aggregate of things, it has no extremity and therefore has no end or limit. It makes no difference in which area of it you take up your position, because, no matter what place anyone may occupy, the infinite extent of the universe in every direction is not diminished. (Lucretius 2001, 1.958-976)

Lucretius, as well as Epicurus, start by explaining what it means to be bounded, limited. If something is bounded, then it has an end, an extremity. But nothing can have an extremity unless there is something beyond to bound it. Lucretius points out that only then there is a point beyond which our vision cannot trace the object. Epicurus is more elliptic and just states that "the extreme point is seen against something else", but what he actually means, as Bailey elaborates, is that we perceive it "as standing out against something which is not 'it'" (Bailey 1926, 184). In short, the limit of a thing is not determined in itself, but by something else outside it. However, in the case of the universe, the all-encompassing whole, the totality of things, there is no such thing outside it, so it does not have an extremity, and thus no end or limit. From whichever point you look and in whichever direction, the universe is extending into infinity. Therefore, the universe is infinite.

Lucretius then introduces the Spear thought experiment, in which he addresses the possibility of a finite universe (1.968-983). He asks us to "suppose that all existing space were finite, and that someone ran forward to the edge of its farthest border and launched a spear into flight". What do we think would happen? "Do you favour the view that the spear, cast with virile vigour, would fly far and reach its target, or do you suppose that something could check it by obstructing its course?" Lucretius insists that these are the only two possibilities ("hypotheses") available to us, and, moreover, whichever we choose, we are ultimately forced to admit that the

universe cannot be finite and is therefore infinite. On the one hand, if the spear continued its flight, then there is something beyond the edge into which it flies. Lucretius does not spell it out, but something beyond would be the void, an empty space that makes movement possible. On the other hand, if the spear bounced back, then there is something beyond the edge that obstructs its flight, namely a body. Either way, there is something beyond—space or some body—, so it must be admitted that the spear "did not start from the end of the universe". Surely, continues Lucretius, you can move the boundary further and further out. Thus the difficulty remains, we are met with the same question and the same two possible answers again and again. Accordingly, says Lucretius, "our conclusion will be that nowhere can a boundary be fixed: no escape will ever be found from the limitless possibility of flight". However, a constantly moving boundary is not a boundary at all. Clearly, the universe without a boundary cannot be finite and is thus exactly the opposite—infinite.

This is a good example of an ancient thought experiment, as characterised by Katarina Ierodiakonou (2011). It is based on a hypothetical scenario that starts with an imaginary assumption—the universe is finite and someone runs towards the very edge of the universe and throws a spear. What is going to happen? It leaves two equally plausible possibilities that we need to investigate open—the spear will continue its flight or it will bounce back obstructed by something that is beyond. In other words, we are faced with a dilemma that is introduced by an imaginary assumption. What is then distinctive of the ancient thought experimenting is the fact that "the imaginary hypothesis initiates a process of thinking without a previously settled or determined conclusion" resulting in a series of clearly stated arguments on the basis of which we "make up our mind on a particular subject" (Ierodiakonou 2011, 48).

When the initial assumption is the opposite of what we want to establish, as it is indeed the case in the Spear thought experiment, then a thought experiment has the structure of a *reductio ad absurdum* argument. That is, the assumption that the universe is finite leads to an absurdity—its boundary cannot be fixed, it is constantly moving outwards, and a boundless universe is certainly not finite (here I am disregarding the possibility of a non-Euclidian space), but exactly the opposite—therefore we reject the assumption and thus prove its opposite. The universe is infinite. Many ancient thought experiments are of this kind.

II. The Spear Thought Experiment: Illustration or Independent Proof?

Certain scholars believe that the Spear thought experiment does not prove something new, but serves as an illustration of the first argument, i.e. the Limit argument (e.g. see Bailey 1926, 184). Thought experiments are often used in such a way, e.g. Einstein uses thought experiments (commonly named "trains/embankments thought experiments") in his popular presentation of the relativity, so that the readers can grasp its essentials without needing to deal with the complicated mathematical apparatus of theoretical physics (see Roux 2011, 12-13). Admittedly, infinity is as difficult to comprehend as relativity, so an illustration would not hurt. However, the Spear thought experiment is not a simplified presentation of an otherwise well-founded theory, its purpose is to disprove the possibility of a finite universe, thus proving that it is infinite. It can be an illustration only in the sense that it attempts to prove the same thing as the Limit argument does—not in general terms, but based on a concrete example, with the help of an image, a narrative. Such presentation is known to help when dealing with difficult, abstract issues. Lucretius could be thinking that a picturesque guise would add credence to the argument, inclining readers to accept it more readily. If we were able to ascertain that Epicurus did not employ it in his account and it is in fact Lucretius' addition, that would strengthen this interpretation. Unfortunately, too little of Epicurus' work survived to come to any conclusion. Scholars do not even agree whether Lucretius only draws from Epicurus' works (only fragments survived from *On Nature*, and the "Letter to Herodotus" preserved in Diogenes Laertius' *Lives of Eminent Philosophers* is a very short summary of natural philosophy) or from other, yet unknown, Epicurean sources (see Warren 2010 and Farrell 2010).

Moreover, Lucretius' thought experiment is very similar to an older thought experiment, which was most likely first devised by the Pythagorean Archytas of Tarentum (first half of the fourth century BCE), in which a man is standing at the very edge of the cosmos and tries to extend his hand or his staff outside. The Stoics also used it in order to support their cosmology, according to which a finite cosmos is situated in an infinite void. Thus, it does not sound improbable that Epicurus and other Epicureans used it too. But again, it could also be the case that Lucretius was unaware of this version and devised his own independently. Furthermore, although we cannot know how much of *De rerum natura* is Lucretius' own contribution and how much he just translated from the Greek source, it would be quite unusual if he had not added anything of his own. He certainly tried to make the material more easily accessible to his Roman readers and it is plausible

to assume that he is at least responsible for the Roman appearance of the Spear thought experiment. It uses military terminology (the Latin "procurrat" refers to the run before throwing the spear and "contortum" to the actual throw of the spear) and may allude to an old Roman custom, according to which the declaration of war was made by a priest throwing a spear from the city boundaries (Ierodiakonou 2011, 40-41).

We can sidestep the authorship issue and consider what the added value of a thought experiment actually is. What can it achieve that an argument cannot? According to John Norton, thought experiments have a rhetorical value, but, epistemically, they are "just ordinary argumentation, disguised in some vivid picturesque or narrative form". Acquired knowledge, the outcome of the thought experiment, is derived from premises which are "introduced explicitly or tacitly into the thought experiment" (Norton 2004, 1139). We might think they are epistemically superior to arguments because they can present the content in a new, unusual way, highlighting previously unnoticed aspects. However, the same outcome is achieved when we construct an argument based on all (explicit and tacit) assumptions of the corresponding thought experiment. (Norton 2004, 1139; see also Norton 1996).

According to Norton, thought experiments are then simply shortcuts. However, others again claim that they are not dressed-up arguments, but epistemically significant, independent devices that can provide information not present in the premises of the corresponding argument. For example, Tamar Szabó Gendler (2010) claims that this is due to a difference in belief formation. Specifically, contemplation of imaginary scenarios involves the manipulation of images, which results in quasi-sensory intuitions providing new beliefs about the world. Gendler calls such intuitions "quasi-sensory" because the processes involved are similar to observation—the difference being that we observe an imaginary scenario instead of the real world. On the other hand, in the evaluation of arguments, the conclusion is inferred from the premises by explicit, logical reasoning. The latter is a successive process, while intuitions just appear instantly in our mind. And since the processes are different, the resulting beliefs can also be different. The thought experimenting has also been explained as mental modelling, where the narrative of a thought experiment guides us in the construction of a mental model corresponding to the depicted situation, which we then manipulate, thus arriving at new beliefs about the world (Nersessian 1992 and 2018; Miščević 1992).

However, regardless of which interpretation of thought experiments we adopt, the Spear thought experiment is not merely an illustration of the Limit argument, but an independent proof. According to one interpretation, every

thought experiment is an independent cognitive device with its own mechanisms and thus different from any argument. According to the other (Norton's argument) interpretation, thought experiments are reducible to arguments, but the argument that corresponds to the Spear thought experiment differs from the Limit argument. The latter shows that the universe is infinite, while the former shows that it cannot be finite and, consequently, must be infinite.

The fact that the Spear thought experiment not only shows that the universe is infinite, but also disproves the rival theory in the process, speaks of its value and why it is worthy of separate consideration. Unlike the Limit argument, it also seems to be less dependent on the Epicurean theory, thus more persuasive for neutral audience. This transpires from the treatment of both proofs, not only Lucretius' versions but also those proposed by other philosophers before him. As mentioned, the Spear thought experiment is similar to the thought experiment that was probably devised by Pythagorean Archytas of Tarentum in first half of the fourth century BCE, while a simpler version of the Limit argument is attributed to the fifth-century Eleatic Melissus and was also used by early atomists (Furley 1987, 137; Avotins 1983, 425). In *Physics* (3.4, 203b15–30; transl. by Hussey 1993), Aristotle lists both amongst the most plausible arguments for the existence of something infinite.

We recognise the Limit argument in the fourth, stating "that what is limited always reaches a limit in relation to something, so that there can be no [ultimate] limits, since one thing must always reach a limit in relation to another" (203b20–21). Aristotle refutes the main idea that the limit of something can only be determined by something else by introducing a conceptual difference between "being limited" and "being in contact". The latter is indeed relative—"everything [in contact] is in contact with something" else, while the former is not, thus not requiring anything else—a thing is limited by reference to itself, nothing else is required. Thus, the fact that something is limited does not imply that there is something outside it, and the conclusion—the sum of all spatially extended things cannot be limited—does not follow. Aristotle further explains that the two concepts get mixed up because some finite things are in contact with something else, but, as he puts it, this is merely an "accident of some finite things" (208a11–14).

Aristotle has a point. It seems possible to explain what a limit is without any reference to something beyond it, but in the Epicurean context we need to consider the peculiar role perceptions play in their theory. Namely, Epicureans, in order to ward off scepticism concerning knowledge, claim that all perceptions are true. Lucretius defends this with the help of an

illustration: as a building will collapse if you use crooked measuring tools in its construction, so will reason if it starts from false perceptions, "your reasoning about things must of necessity be distorted and false if the senses upon which it is based are themselves false" (Lucretius 2001, 4.513–521). Elizabeth Asmis (2009) summarises their methodology as follows. First, they formulate problems with the help of initial concepts, and then infer what is not observed with the help of perceptions and feelings. Initial concepts themselves are derived from sensory observations, which makes them evident; they are "seen". Since an initial concept is "grasped prior" to the pursuit of an inquiry, Epicurus coined the special term "prolepsis" (preconception) to signify it. And these preconceptions replace definitions. They vary in their complexity. The most basic are the preconceptions of sensory qualities (red, round, sweet), then those of individuals (Plato), and of general kind (human being, cow, body). The most complex ones are those of even more general kind (goodness of a poem, justice, cause, god). They are formed by the mind that gathers similar appearances (direct objects of perceptions that an influx of atoms produces in us). In case of the preconceptions of individuals, the mind gathers a selected number of perceived similarities into a single conception; in the case of the most complex preconceptions, it draws connections involving several preconceptions.

According to this account, Lucretius and Epicurus arrive at the concept of a bounded body, employed in the Limit argument, by gathering general features of many perceived bounded bodies. In fact, they both justify the requirement of something external bounding a finite thing by how we perceive it or "see" it. Lucretius points out that only if there is a point beyond which our vision cannot trace a thing do we perceive a thing as being bounded. Furthermore, Epicurus similarly states that we perceive a boundary as standing out against something else. Therefore, neither would accept Aristotle's claim that a limit of a body can be set without reference to something else.

Aristotle actually did not provide much justification for the distinction between being in contact and being limited. He simply states that these are two different things and explains why one would conflate them by the fact that sometimes finite things are also in contact with other things, thus exhibiting an accidental property of being in contact with something else. Aristotle, however, forgets to mention here how often it happens that finite objects are in contact with something else. It follows, from his rejection of the void and his definition of place as "the first unchangeable limit of that which surrounds" (212a20), that every limited thing in the cosmos is in contact with some other thing all the time. Thus, the only exception is the cosmos, "the heavens as a whole", which according to Aristotle is not

surrounded by anything. But that means that cosmos is not in any place, it is actually nowhere, and Aristotle is at pains to explain this puzzling consequence of his theory (Aristotle 1993, 212b8–21). Commentators struggle with it, as it is already unclear what Aristotle meant with the *ouranos* in the relevant passage: outer sphere of the heavens, the heavens as whole, or the cosmos as a whole? But there are other difficulties. Keimpe Algra thus suggests that the issue of the emplacement of the heavens is unsolvable within Aristotle's theory (2019, 33–36).

Thus, from the Epicurean point of view, Aristotle's distinction fails since it is not based in experience and is hence no threat to their Limit argument. They can insist that a thing is limited by something else, since this is how they see limited things. However, Aristotle's insistence on the exceptional status of the cosmos, or the universe, is not entirely misguided. Often the whole is more than just its parts: in particular, it is not necessary that the common property of all parts is also a property of the whole. Or it could be said that things are in contact with each other exactly because they constitute a whole, while they are limited in themselves, without reference to surrounding things. It just happens that we discern the limit in contrast with something else, but this is a fact about us and not about the nature of a limited thing. We should not forget that Epicurean theory is itself problematic. Every theory that relies heavily on experience needs to justify generalisations, and Epicureans must additionally establish that the appearances, which are direct objects of perception, are really caused by something "outside"; namely, that we perceive what exists objectively. They also need to explain how we know things that are imperceptible, "not-apparent". The infinity of the universe undoubtedly counts as such. For example, Lucretius ends the argument with the claim that an observer would, from any position in the universe, see that it extends in all directions, never diminishing (1.965-967). An observer clearly cannot see this; we cannot see infinity. Our vision is limited and the reasonable conclusion based only on observation is that we live in a finite system bounded by the sphere of stars. Clearly, the dispute between the proponents of the finite cosmos and those of the infinite universe is not based solely on perception but also on speculation.

Epicureans do have an idea about how to deal with proving what cannot be observed, what is "non-apparent". Indeed, we cannot use the method of "witnessing", where a thesis is proved by direct testimony of senses, but we can use the method of "no counterwitnessing", where a thesis is proved if a thorough investigation of the phenomena (what appears directly in sensory awareness) does not reveal anything that contradicts it (Asmis 2009, 99). Thus, in the case of the infinite universe, a thorough examination of the

phenomena should not reveal anything that contradicts its being unbounded and infinite.

However, it is easier to falsify than to verify something, and here lies the advantage of the Spear thought experiment over the Limit argument. It falsifies the opposite thesis that the universe is finite and it confirms that it is infinite. In a similar vein, for instance, Lucretius proves that nothing comes to be from nonbeing (Lucretius 2001, 1.159-173). If this were not the case, Lucretius suggests, "nothing would need a seed". This would result in a world in which "human beings could spring from the sea, squamous fish from the ground, and birds could be hatched from the sky". This clearly is not the case, we do not observe such happenings, and because the phenomena contradict the assumption, Lucretius rejects it and confirms that nothing comes to be from nonbeing. The issue of the extent of the universe is different though, because it is not clear how it would affect the phenomena if it had been finite as opposed to infinite. Well, except that in the first case it would have an edge and in the second it would not have. Clearly not something we could observe, so Lucretius turns to the next best thing—imagination. He invites us to imagine someone running towards the edge of the farthest border of the existing space and, based on this example, shows that, given our preconceptions concerning space, this is an impossible proposition. We simply cannot fix the edge of the universe, it is constantly being moved further and further out, thus making us realise that there can be no such edge and that the universe can only be infinite. The finitude of the universe is thus counterwitnessed by our preconceptions concerning space, while its infinity is not. It is consistent with common perceptions and preconceptions.

The biggest advantage of the Spear thought experiment is that it does not depend on the tenets of the Epicurean theory as the Limit argument does. For example, it does not explicitly rely on the preconception of limit, but it rather invokes our ordinary intuitions about space and about how it works. It is all in all a persuasive and strong proof of the infinite universe. This is also attested by the popularity of Archytas' older version with a man at the outermost edge of heaven extending his hand or a staff into what is outside. It was adopted by the Stoics, and also employed by early modern thinkers such as Giordano Bruno, Pierre Gassendi, John Locke and even Isaac Newton (Ierodiakonou 2011, 38). It is very persuasive. As Carl Huffman emphasises in his detailed analysis of Archytas' thought experiment, experience tells us that Archytas is right. If there is no physical obstacle in front of us, we can extend our hand, so why it would be any different at the edge of heaven? Thus, the burden of proof lies firmly on those who claim that it is indeed different, and they owe us an explanation as to why "at the

edge of the cosmos all of our previous experience and intuitions about space are invalid" (Huffman 2005, 549-550). The same considerations apply to Lucretius' Spear thought experiment.

Aristotle also deems arguments, based on imagination, to be very strong. He summarises them in the fifth argument for the infinity of the universe, "Above all, and most decisively, the argument which makes a common difficulty for all thinkers: because they do not give out in thought, number and mathematical magnitudes and what is outside the heavens all are thought to be infinite" (Aristotle 1993, 203b22-25). Imagination can always imagine something beyond or, more in line with Aristotle's argumentation, thought can always conceive of something beyond when it comes to numbers, geometrical figures and the extent of the cosmos, causing difficulties to everybody who denies their infinity. By the third example, Aristotle probably means Archytas' thought experiment; Simplicius, in his commentary on Aristotle's *Physics* (467.26–468.3), tellingly adds it immediately after his discussion of this fifth argument (Huffman 2005, 545–546).

Aristotle (208a14–19) ultimately rejects such imaginative arguments by pointing out that us being able to conceive or imagine something in a certain way does not really make it to be that way, for example, "one might think of each of us as being many times as large as himself, increasing each of us *ad infinitum*; but it is not for *this* reason, because someone thinks it so, that anyone exceeds this particular size that we have, but because it is the case" (208a14–19). True, our ability to conceive something does not make it real, or even possible, and, clearly, if we had something like this in mind, then it would be "absurd to rely on thought", as Aristotle purports. However, the role of conceivability, or of imaginability, is epistemological, not metaphysical; it guides us in finding out what is possible, impossible, and sometimes also what is real. Aristotle is hence more plausibly denying that conceivability can be a method of acquiring knowledge of reality. Indeed, it is a fallible method, but that does not automatically mean that it is worthless, as Aristotle would have it. He does not offer any further support for his extreme claim. Simplicius clearly did not think Archytas' proof is irrelevant—after its description, he suggests that "perhaps this argument will keenly trouble us too who say that there is nothing outside the heaven" (Huffman 2005, 541). Other Aristotelian commentators from the late Antiquity took it and the Stoic version seriously as well.

To sum up, Lucretius' Spear thought experiment makes a persuasive, strong case for the infinity of the universe. It addresses our ordinary intuitions about space, thus making it difficult to avoid its conclusion by seeking refuge in theoretical considerations. Non-Euclidian geometry

taught us to conceptualise space as being finite and unbounded, but we cannot fault Lucretius for not considering it. An interesting question, worthy of investigation, is how this physical concept of space meshes with our ordinary intuitions about space and what it means for the use of thought experiments in general. However, this goes beyond the scope of this article.

III. Is it a Thought Experiment?

At the end let me shortly address the concern that it is illegitimate to speak of ancient thought experiments. They certainly did not call them so at the time and did not, at least not explicitly, distinguish them from ordinary arguments. In *The Routledge Companion to Thought Experiments* (2018), all three chapters dealing with ancient thought experiments (Ierodiakonou; Becker; Corcilius) mention and ultimately dismiss this concern. Although the ancients did not have a technical term for "thought experiment" and did not think of them as a distinct tool of investigation, they did use them extensively. They devised hypothetical scenarios in order to prove or disprove certain theories just as contemporaries do. Ierodiakonou's account (2011) of ancient thought experiments described above shows that they are in most respects similar to today's thought experiments. It sounds strange not to count them as thought experiments simply because the ancients did not have a specific name for them. As it would be strange to say that someone who just hammered a nail into a plank did not use a hammer to do so because they do not know that the tool they used is called a "hammer".

The ancients used thought experiments successfully and, moreover, seemed to recognise their difference with respect to other arguments they used to prove or disprove theories. Namely, in these cases, a generally applicable conclusion is drawn from the consideration of a specific example. As Ierodiakonou mentions (2018, 33), typical thought experiments are, in ancient texts, often referred to as "paradeigmata", i.e. as "examples"; the term "paradeigma" literally means "the sample, pattern plan, model, sketch that is placed besides something else for comparison or juxtaposition". We are therefore justified in speaking of the Spear thought experiment since it has all the characteristics of a thought experiment, even though Lucretius does not call it so.

There is also a more specific concern stemming from the historical fact that the study of thought experiments started with the scientific ones that played an important role in the birth of empirical natural sciences in the seventeenth century, e.g. Galileo's thought experiments. Pierre Duhem, Ernst Mach, and Alexius Meinong focus on thought experiments in physics and model them on actual experiments, the main difference being that the

latter are not conducted in thought, but in the real world. From this it follows that the characterisation of a thought experiment depends on the notion of experimentation as it was established in modern natural sciences. In "Aristotle and Thought Experiments" (2018), Klaus Corcilius characterises experimentation as "an artificial means to make available to repeatable experience hitherto unobservable facts about nature" (Corcilius 2018, 57). But the question is then how can such a conception of a thought experiment be applied to the periods before such understanding of experimentation was developed. And the ancients do not really use, and certainly not systematically, experiments in order to test their theories, but rely on immediate experience instead (for a more nuanced account, see Lloyd 1991).

This concern is unfounded. First, if we accept such a close link between thought and actual experiments, we have to exclude the majority of contemporary philosophical thought experiments as well, which is quite extreme. However, as I mentioned above, an emphasis on this link is just a historical accident and not the expression of the true nature of thought experiments. Once investigation expanded to other areas, especially to philosophy, the "experimental" aspect of thought experiments lost some of its importance. Today, a thought experiment is characterised as a device that involves contemplating an imaginary scenario with the aim of acquiring a piece of information, while further characteristics depend on the area in which the thought experiment is being conducted. In certain areas, the experimental aspect is more relevant, in others less. Surely, in philosophy, which is not known for its experimental methods, the link with experimentation is not that important. It could be recognised in the fact that an imaginative scenario is like an idealised model on which we test a theory, its consequences, and how they mesh with other theoretical consideration and our intuitions. And the Spear thought experiment belongs to ancient cosmology, a speculative discipline that is part of philosophy, so it is more appropriately classified as a philosophical and not as a physical thought experiment.

Second, we should not forget that this thought experiment has a long history. After antiquity, we encounter it again in the Middle Ages (e.g. Thomas Aquinas, John Buridan, Nicholas Oresme), but, crucially, in modern historical periods as well. It is used by philosophers (e.g. Giordano Bruno, John Locke), but, even more important for the present discussion, by physicists (e.g. Pierre Gassendi, Isaac Newton) as well. This puts us in a strange position. If we allow the label "thought experiment" only for the examples employed in modern times, we end up with quite an arbitrary classification of what counts as a thought experiment and what does not. For example, in what way is Gassendi's or Newton's use of the example

different from Archytas' or Lucretius'? And how would we classify Bruno's? In certain ways he belongs to modern times, while in others he does not. He clearly rejects the mathematisation of nature. Therefore, I conclude that the proposed restriction on what counts as a thought experiment is impractical and unfounded. The Spear thought experiment is rightly called so.

Bibliography

Algra, Keimpe. 2019. "Aristotle's Account of Place in *Physics* 4: Some Puzzles and Some Reactions". In *Space, Imagination and the Cosmos from Antiquity to the Early Modern Period*, edited by Frederik A. Bakker, Delphine Bellis and Carla Rita Palmerino, 11-39. Dordrecht: Springer.

Aristotle. 1993. *Physics, Books III and IV*, translated by Edward Hussey. Oxford: Clarendon Press.

Asmis, Elizabeth. 2009. "Epicurean Empiricism". In *The Cambridge Companion to Epicureanism*, edited by James Warren, 84-104. Cambridge: Cambridge University Press.

Avotins, Ivars. 1983. "On Some Epicurean and Lucretian Arguments for the Infinity of the Universe". *The Classical Quarterly*, Vol. 33, No. 2: 421-427.

Bailey, Cyril. 1926. *Epicurus: The Extant Remains*. Oxford: Clarendon Press.

Becker, Alexander. 2018. "Thought Experiments in Plato". In *The Routledge Companion to Thought Experiments*, edited by James Robert Brown, Yiftach Fehige and Michael T. Stuart, 44-56. London and New York: Routledge.

Corcilius, Klaus. 2018. "Aristotle and Thought Experiments". In *The Routledge Companion to Thought Experiments*, edited by James Robert Brown, Yiftach Fehige and Michael T. Stuart, 57–76. London and New York: Routledge.

Farrell, Joseph. 2010. "Lucretian Architecture: The Structure and Argument of the *De rerum natura*". In *The Cambridge Companion to Lucretius*, edited by Stuart Gillespie and Philip Hardie, 76-91. New York: Cambridge University Press.

Furley, David. 1987. The Greek Cosmologists, Volume 1: The Formation of the Atomic Theory and Its Earliest Critics. Cambridge: Cambridge University Press.

Gendler, Tamar Szabó. 2010. "Thought Experiments Rethought—and Reperceived". In *Intuition, Imagination, and Philosophical Methodology*, 42-52. New York: Oxford University Press.

Huffman, Carl A. 2005. Archytas of Tarantum: Pythagorean, Philosopher and Mathematician King. Cambridge: Cambridge University Press.

Ierodiakonou, Katerina. 2011. "Remarks on the History of an Ancient Thought Experiment". In *Thought Experiments in Methodological and Historical Contexts*, edited by Katerina Ierodiakonou and Sophie Roux, 37-49. Leiden: Brill.

Ierodiakonou, Katerina. 2018. “The Triple Life of Ancient Thought Experiments”. In *The Routledge Companion to Thought Experiments*, edited by James Robert Brown, Yiftach Fehige and Michael T. Stuart, 31-43. London and New York: Routledge.

Lucretius Carus, Titus. 2001. *On the Nature of Things* [*De rerum natura*], translated by Martin Ferguson Smith. Indianapolis/Cambridge: Hackett Publishing Company, Inc.

Lloyd, Geoffrey Ernest Richard. 1991. “Experiment in early Greek Philosophy and Medicine”. In *Methods and Problems in Greek Science: Selected Papers*, 70-99. Cambridge: Cambridge University Press.

Miščević, Nenad. 1992. “Mental Models and Thought Experiments”. *International Studies in the Philosophy of Science*, Vol. 6, No. 3: 215-226.

Nersessian, Nancy J. 1992. “In the Theoretician’s Laboratory: Thought Experimenting as Mental Modeling”. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2: 291-301.

Nersessian, Nancy J. 2018. “Cognitive Science, Mental Modeling, and Thought Experiments”. In *The Routledge Companion to Thought Experiments*, edited by James Robert Brown, Yiftach Fehige and Michael T. Stuart, 309-326. London and New York: Routledge.

Norton, John D. 1996. “Are Thought Experiments Just What You Thought?”. *Canadian Journal of Philosophy*, Vol. 26, No. 3: 333–366.

Norton, John D. 2004. “On Thought Experiments: Is There More to the Argument?”. *Philosophy of Science*, Vol. 71: 1139-1151.

Roux, Sophie. 2011. “Introduction: The Emergence of the Notion of Thought Experiments”. In *Thought Experiments in Methodological and Historical Contexts*, edited by Katerina Ierodiakonou and Sophie Roux, 1-33. Leiden: Brill.

Warren, James. 2010. “Lucretius and Greek Philosophy”. In *The Cambridge Companion to Lucretius*, edited by Stuart Gillespie and Philip Hardie, 19-32. New York: Cambridge University Press.

Chapter Six

Intoxication, Death, and the Escape from Dialectic in Seneca's EM[1]

David Merry

Introduction

In this essay, I wish to return to a topic about which there has been a quiet but sustained discussion in the quiet but sustained field of Seneca scholarship; namely, Seneca's criticism of dialectic in the *Epistulae Morales.* I will focus in particular on the critique as it is developed in *EM* 82, 83, and 87, as this is where Seneca presents his ideas on dialectic in their fullest and most developed version.

Seneca's criticism of dialectic has not secured for him much admiration among contemporary commentators. Cooper, for instance, argues in line with Hadot (1969), that Seneca's treatment of dialectic reflects a "preference for rhetorical appeals to a person's feelings over solid reasons why the conclusions really are true" (Cooper 2004, 320), in contrast to "the original Stoics, who were firm and clear about the far greater value of sound and solid reasoning for establishing such conclusions" (Cooper 2004, 320). In a similar vein, Malcolm Schofield (1983) argues that Seneca's critical approach to dialectic reflected his misunderstanding of the practice and concomitant underestimation of its philosophical fertility.

Inwood (2007) takes a more apologetic approach. Following Barnes (1997), he argues that Seneca's criticism of dialectic is aimed

> "/…/ at its excesses, and not as the practice as such". Inwood stresses that Seneca is an ironic author, and so we ought not to take his criticism of the

[1] This article grew out of a chapter of my PhD thesis, and I have discussed many of the ideas with various people. I am particularly grateful for comments from Stephen Menn, Jonathan Beere, Gisela Striker, Richard Kraut, Johanna Schmitt, and audiences in Berlin and Maribor.

> practice at face value: after all, Seneca spends a lot of time discussing dialectical technicalities, which is surprising for an author who wishes to dismiss them: "silence would have been more effective". (Inwood 2005, 218)

Furthermore, Inwood argues that, in preferring vivid exemplification to formidable argument, Seneca is committing no greater philosophical sin than preferring Posidonius to Chrysippus, or than situating himself in line with the Aristonian idea that ethics is the only truly necessary branch of philosophy.

My purpose in this essay is to argue that Seneca's target in *EM* 82 and *EM* 83 is not dialectic as such, but rather refutation, and that he raises valid and philosophically interesting concerns about the practice of refutation in philosophy, concerns that have a continuing relevance to how we ought to understand the relationship between thought experiments and argumentation in philosophy.

I. Seneca's Critique of Dialectic in EM 82 and 83

Dialectic was one of the central argumentative practices of Plato's academy, and continued to be an important aspect of the philosophical life both in the Lyceum and in the Stoa. The term *dialectic* is sometimes also used to mean an ability in such argumentative practices, but, as we will soon see, Seneca's target is the practice itself, rather than the ability (Kakkuri-Knuutilla 2005). The practice of dialectic is recorded most vividly in Plato's dialogues. At times, especially the *Gorgias,* the *Euthydemus,* and the *Sophist,* Plato was careful to differentiate the practice from sophistic and rhetorical practice. Dialectic is differentiated from rhetoric in that it involves a dialogue between two people, using short speeches and questions and answers. Precisely what differentiates dialectic from sophistic is harder to say: nevertheless, Aristotle defines sophistic as dealing in merely apparent refutations, whereas dialectic is the art of real refutations. (*SE*, 164a20)

Aristotle's *Topics* is a handbook for students to succeed in dialectical encounters, and it also stresses the interpersonal nature of dialectic. (Smith 1997 and 1991; Slomkowski 1997) Dialectic, unlike philosophy, is necessarily carried out with another person; the method developed in *Topics* is one of asking questions and answering them. (*Topics* 1.1.100a18-20, 8.1.155b1-15) The Stoics give two different definitions of dialectic: it is "the science of correct discussion in regard to discourses conducted by question and answer, so that they also define it as the science of what is true and false and neither." (Diogenes Laertius 7.41--4, Long and Sedley 31A, trans. Long and Sedley) Understood as the science of what is true, false, and neither, Stoic

dialectic involved the development of a sophisticated propositional logic widely seen as an important advance over Aristotle's syllogistic. However, the Stoics were also interested in the dynamic nature of dialectical arguments: one Stoic solution to the Sorites, for instance, involves strategically refusing to answer questions (Williamson, 1994). For an extended and fascinating discussion, see Castagnoli's "How Dialectical was Stoic Dialectic?" (2010).

There is a lot that might be said in favour of dialectic as a philosophical practice. Short questions and answers allow the answerer's commitments to be tracked and negotiated in an ongoing way; the answerer can, and often would, disambiguate questions. Dialectical answerers are held accountable for the consistency of their commitments; a skilled questioner might lead an answerer into refutation, a distressing state which would reveal to the answerer a gap in their understanding. Dialectic, unlike rhetoric, minimises appeal to emotions, focusing on the logical connection between proposition and proposition. Nevertheless, as Plato's interlocutors sometimes complained, dialectic was a *social* practice: Polus argued that Gorgias was refuted by Socrates because he was ashamed to say what he truly thought, (*Gorgias*, 461b-d) and Aristotle would later write that one must consider one's reputation in making key decisions about what propositions to defend in dialectic (*Topics,* VIII.9.22-9). The philosophical value of this social aspect of dialectic will certainly be more controversial: one might see philosophy as a practice of negotiating norms within a community, in which case the presence of social pressures in shaping argumentation will be a strength of the practice; (Duncombe and Dutilh Novaes, 2016) one might also see it as a practice of individual freedom against social norms, in which case these will be a shortcoming. In the *EM,* Seneca makes several sustained criticisms of dialectic as a social practice. In *EM* 45, for instance, Seneca takes aim at the practice of studying fallacies, a practice he sees as largely idle and useless for obtaining philosophical knowledge. The subject of this essay is his critique of dialectic in *EM* 82 and 83, where he will argue against the use of dialectic in turning people towards virtue, and in producing moral knowledge, which for a Stoic are, after all, ultimately the same.

Seneca's targets in *EM* 82 and *EM* 83 are, more specifically, two syllogisms by Zeno of Citium. The first, in *EM* 82, is about death, and the second, in *EM* 83, about drunkenness. About drunkenness, Zeno argued: "No one entrusts a secret to one who is drunk. But one does entrust a secret to a good man. Therefore a good man will not be drunk" (Seneca 2015, 279).[2] About death, Zeno argued: "Nothing bad is glorious; but death is glorious; therefore death is not bad." Seneca records a relatively adroit

[2] Translations from Graver and Long.

philosophical back-and-forth about these syllogisms: parallels, arguments with the same structure seeking to show their invalidity were propounded: you would not trust a secret to a sleeping person either, so by this logic the wise man does not sleep; what is glorious is also not indifferent, so death is not an indifferent. As we will see, Seneca thinks that both arguments are flawed, but that in neither case does the parallel identify the core problem with the argument, and that the whole discussion of these parallels was a waste of time.

Seneca is more sympathetic to the argument about death than the argument about drunkenness, so we shall deal with that one first. Seneca thinks the person who created the parallel argument simply misunderstood the Stoic doctrine of the indifferents. As Seneca understands this doctrine, indifferents are sometimes good and sometimes bad, depending on whether they are associated with virtue or vice: "all such things are not in themselves either honorable or glorious, but any of them that virtue meets and handles is made honourable and glorious by it" (*EM* 83.12). So indifferents are things that are sometimes good, and sometimes bad. Death is not glorious in itself, but it is sometimes glorious, and that is enough to show that it is not bad: truly bad things, such as vices, never become glorious because of being handled by virtue. So the parallel argument does not follow, because it is simply not true that what is glorious is not an indifferent.

It is worth stopping a moment here, as what Seneca says in these passages about preferred indifferents is rather striking. As contemporary commentators are aware, the Stoics certainly did see the indifferents as lacking value, and, among other arguments, followed Socrates' reasoning that both good and bad use can be made of preferred and dispreferred indifferents (Barney, 2003). Yet the Stoics, aside from Aristo, thought that being a preferred indifferent was, at least in normal cases, a *reason for action* (Barney, 2003), and indeed Cicero, probably following Carneades, would use the claim that indifferents were reasons for action to argue that the Stoic position was ultimately untenable: if indifferents offer reasons for action, they must have value, and if they do not, they must not (*De Finibus*, 4.17). Yet here Seneca is not only saying that indifferents do have value (albeit extrinsic value), but that this value changes so that at times it goes against the valence of the indifferent: preferred indifferents sometimes have negative value, and dispreferred indifferents sometimes have positive value. This makes it even more difficult to see how the valence of an indifferent can act as a criterion of action, for if I, as a virtuous person, select a dispreferred indifferent, it will become good in my hands: what possible reason can I have then for rejecting it? An already difficult point for the Stoics has become sharper.

Klein (2016) offers a neat solution. According to Klein's reading, indifferents act as reasons for action, not because they have value, but as epistemic reasons. Providence inclines us towards preferred indifferents, and away from dispreferred indifferents, because in most cases this will lead us to actions in line with the divine plan. When a course of action is unclear, selecting preferred indifferents and avoiding dispreferred indifferents will then be a good way to proceed in terms of doing the appropriate thing. However, this is *not* because the indifferents have value, but rather just because this is more likely to accord with the divine plan. On Klein's reading, there is plenty of space for indifferents to change in value as they become united with virtue or with vice, as there is no need to accord them any value—be it quantitatively or qualitatively different from the value of virtue—in order to explain how they can be reasons for action.

Returning to the argument above: Seneca's appraisal is that it is sound: it has true premises and a valid inference. Yet he still thinks that Stoics should not use this argument, certainly not in convincing people not to fear death: he thinks that it is important we "not reduce these matters to the rules of dialectic and those tired old conundrums of professional logicians", that "all that sort of thing ought to be thrown out" (*EM 82.10*). If Seneca's position were simply that the argument is insufficiently persuasive, then he would, at this point, have been ceasing to do what we now recognise as philosophy, and Cooper's criticism, that he does not seem much concerned with the truth, would be quite a valid one. When we are practicing philosophy, we seek to offer people good reasons; while we may do our best also to be persuasive, ultimately a philosophical audience has responsibilities as well. If they stubbornly refuse to change their minds even after conceding that an argument is good, then we ought not to start offering them a marketing campaign.

Seneca does repeatedly complain that the argument is not persuasive, but his diagnosis of why it is unpersuasive is essential to understanding the philosophical interest in his criticism (*EM* 82.9, 19, 21). The key point is that the interlocutor feels "cheated", and will be "saying [things] different from what he actually believes" (*EM 82.19)*. Cooper takes this to mean that the interlocutor feels cheated only when the conclusion is drawn, that this would be the only point at which there is a mismatch between what the interlocutor says and what they believe. However, if we consider Seneca's preferred therapy for the fear of death, we can see that he must have thought there was a serious issue with the premise "death is glorious".

In *EM* 24, Seneca discusses how to deal with fear. The process he describes is that of *praemeditatio*, he asks his reader to: "take up each of these things in turn, and summon to mind those who have thought little of

them" (*EM* 24.3) In the letter, he provides several examples, including that of Cato's death. Closer to hand, in *EM* 82, we find the example of Cato again, and his glorious death, compared with that of the disgraceful death of Decimus Brutus. The point of providing these examples is ultimately to understand that death is an indifferent (Armisen-Marchetti, 2008). The practice does this by offering reminders of the truth of the claim that death's value changes depending on whether it is associated with virtue or vice: Seneca thinks that we should focus our efforts on this. Finally, later in *EM* 82, we find Seneca recommending the approach of Leonidas and a Roman commander, who offered themselves as examples of people who treated death lightly, and so gave an image of the greatness of spirit that Seneca sees as so appealing (*EM* 82.21-22). These cases are aimed not only at strengthening the conviction that death is not bad, but also at strengthening or bolstering conviction that some deaths are glorious.

Returning, then, to the problem of Zeno's argument. The audience might assert that "death is glorious", they might think it, but they lack conviction. The problem of the argument, then, is not simply that it is not persuasive, but that the argument is not from common ground. It may be going too far to say that the audience does not believe this statement, but their conviction in this claim is unlikely to be very deep. However, in a dialectical encounter, one has to say either yes or no, and there would be a social cost to denying that death was glorious, or even to expressing reservations about this claim. The sense of being trapped by Zeno's argument, then, comes from the social pressure to publicly assert that death is glorious, even if one secretly doubts it. The interlocutor is then stuck in an uncomfortable position, where they want to find a way to assert both that death is glorious and avoid asserting that death is bad, and this leads them to look for logical loopholes, as with the parallel argument (*EM* 82.19).

There is, on the basis of the evidence I have presented so far, still considerable room to doubt my analysis of Seneca's discomfort with Zeno's argument. Certainly, exploiting the gap between what one truly believes, what one thinks, and what one is willing to assert are important considerations for the evaluation of the argument from a dialectical point of view, and from the evidence I've presented Seneca probably did think these came apart in the case of Zeno's argument. Further, Seneca is quite clear in pointing out that the problems with this argument relate to the feeling of being trapped and the social pressure of the practice of dialectic. Nonetheless, Seneca never explicitly says that the interlocutor doubts that death is glorious, and we may have expected him to be clearer.

The case for the argument about death is considerably bolstered if we consider Seneca's diagnosis of the argument about drunkenness. Seneca's

analysis of the argument about drunkenness is different in several details from the argument about death, but he sees the problem as having the same basic cause: the use of dialectical pressure. And in the case of the argument about drunkenness, Seneca is quite clear that the dialectical pressure being exerted is on the interlocutors to accept a premise they do not believe.

The premise in question is "no one entrusts a secret to one who is drunk". Seneca is unimpressed with Posidonius' attempt to save the argument from a parallel by arguing that Zeno's meaning was that "you would not trust a secret to a drunkard". Seneca's response is simply to construct a list of trustworthy drunkards: Tillius Climber, entrusted with the plot to assassinate Caesar, Lucius Piso, a trusted advisor to Augustus and Tiberius, and Cossus, another trusted advisor of Tiberius (*EM* 83.12-13). As Seneca says, "each of us could name people who cannot be trusted with wine, but can be trusted with a secret" (*EM* 82.13). However, just as in the example above, there is a disconnect between what is said and what is thought or truly believed. Zeno's argument is an effective dialectical weapon because people will avoid saying that they would trust alcoholics, or even people currently drunk, for fear of being thought foolish, or even simply out of habit—after all, Seneca does describe this is a platitude. The argument's strength, then, depends on people's willingness to assert things they do not truly believe.

We find, then, a philosophically robust criticism of both arguments in Seneca. The criticism is not on the level of truth and logic, but on another important philosophical level, the level of rational argumentation. For Seneca, the understanding of arguing from common ground in dialectic has serious shortcomings, and encourages arguments that connect with what people are willing to say rather than with what they really believe. A failure to argue from common ground is certainly a philosophical failure. Far from misunderstanding dialectic, Seneca understood the practice all too well.

II. An Alternative to Refutation in Seneca's 87th Letter

In this section I will offer an analysis of the syllogisms in *EM* 87. *EM* 87 is ostensibly about wealth, but in fact wealth is used to exemplify a number of Stoic syllogisms that can be used more generally to argue that an apparent good is not actually good. Barnes (1998), Inwood (2005), and Cooper (2004) all read Seneca's attitude to the Stoic syllogisms in *EM* 87 as similar to his attitude to Zeno's in *EM* 82 and 83, but I will argue that this is incorrect: Seneca actually thinks the syllogisms in *EM* 87 are good arguments, and uses *EM* 87 to illustrate a different approach to common ground.

Seneca's attitude towards the syllogisms in *EM* 87 is different from that of his attitude towards those in *EM* 82 and *EM* 83, although there are several features of the letter that suggest a similar attitude. These features are that the letter closes with a set of similar criticisms: Seneca calls for clearer and more direct argumentation, the same terms he used in *EM* 82 and *EM* 83, and points to a rhetorical context (that of making a speech to the senate) to highlight argumentative shortcomings, as he did by appealing to the rhetorical context of encouraging soldiers at *EM* 82.21-22. However, from what we saw earlier, Seneca's attitude towards Zeno's syllogisms was extremely negative: such syllogisms are dialectical traps and dishonest, and should therefore never be used. Yet, as Inwood (2007) notes, Seneca portrays the sage as deploying several of the syllogisms in *EM* 87 at *De Vita Beata* 24.5. We do not find the sage deploying Zeno's syllogisms about death or drunkenness. Further, the structure of *EM* 87 suggests a more limited reading of the criticisms with which the letter closes: the final syllogism Seneca considers clearly involves a dialectical trap similar to those of Zeno's, this time by playing on an ambiguity.

The final syllogism in the letter is as follows: wealth is made up of many poverties, what is composed of bad things is not good, so wealth is not good. Inwood (2007) mentions a suggestion in conversation by Irwin that this may be a fallacy of composition. My guess is that this is not quite right, but rather that in dialectic, one can secure the idea that wealth is made up of many poverties by asking whether having a small amount of money is considered poverty, and then asking whether wealth is having a small amount of money many times over. Then wealth is entirely made up of bad things. This would not be a fallacy of composition, precisely, because it does not infer generalising about the whole from one part: the issue is rather the incorrect definition of poverty. This is in line with the solution that Seneca alludes to: "the word poverty refers not to possession but to subtraction" (*EM 87.39)*. If we take the criticisms that *EM* 87 closes with as only applying to this final syllogism, then *EM* 87 actually presents a contrast: the earlier syllogisms in *EM* 87 are syllogisms that Seneca sees as sagacious ways to argue, in contrast to the syllogisms in *EM* 82, 83, and the final syllogism in *EM* 87. If this is correct, then we need to develop an account of the features of the syllogisms in *EM* 87 that make them more appealing to Seneca, and that is what I will aim to do in this section, by offering an analysis of two of the syllogisms. The first syllogism I will consider is the second of the letter, which I will set out and then discuss:

> That which can belong to the vilest and most despicable kinds of people is not a good;
> but wealth can belong to the pimp and the manager of gladiators.

Therefore, wealth is not a good. (*EM 87.15*)

The statement of the argument is clear. To prove that something is not good, we simply have to argue that it can be possessed by the worst people imaginable. At a glance, we can see that if this principle is accepted, then it will speak against the goodness of quite a wide range of things generally thought to be good: an excellent reputation, popularity, a family, being loved, pleasure, and of course wealth. We might further see that a Stoic will accept the major premise, because after all a Stoic already believes that virtue is the only good, and does not think that a bad person will have virtues. But, until we have become convinced of this doctrine, it seems unlikely that we will accept the major premise, as there are so many apparent counter-examples to the principle that the worst people cannot have any good things.

From a dialectical perspective, then, this syllogism is remarkably different from the syllogisms of Zeno discussed earlier. Zeno's syllogisms sought to argue from premises that interlocutors would find it difficult to disagree with. This syllogism takes a rather different, and from a dialectical perspective, much less powerful approach: it presents a premise that the interlocutor is certain to reject. Above, I wrote that Seneca was well aware of the need to argue from common ground, and indeed criticises Zeno for failing to do so adequately. When this syllogism is seen in isolation of Seneca's discussion of it, however, it seems that Seneca has simply given up on any project of finding common ground from which to argue. Inwood (2007) points out that the syllogism has the structure of a Peripatetic syllogism—namely *Celarent*—and that this reflects an understanding of a need to engage with one's interlocutors on their terms; be that as it may, there seems little point in using a Peripatetic form if one is going to argue from premises that no Peripatetic would accept. One thing we can say on considering this syllogism is that Seneca is staying true to his rejection of refutation.

Seneca has not, for this, turned his back on the idea that it is important to look for common ground. In his discussion, Seneca considers a Peripatetic rejoinder: "for in the teaching of literature, in medicine, and in navigation, we see that the relevant goods belong to the humblest kinds of people" (Seneca 2015, 303). Here, as Inwood (2007) notes, the Peripatetics are developing the craft analogy: virtue is an excellence of living, and so we should expect its relationship to goods to be the same as those in other skills. In general, skills point towards an end, but are separable from the end, and the end can be held by people without the skill.

This is a natural thought, and Seneca uses it to begin an exercise in considering how virtue and vice are different from skill and ignorance, one that connects intimately with the syllogism he just presented, and which finds common ground with the Peripatetic. Seneca still does not give a refutation, of course, as he thinks refutation is counter-productive, but rather, as in *EM* 82 and 83, turns to something that is more akin to a thought experiment for which the syllogism can be considered a kind of blueprint.

First, Seneca makes a point about how virtuous people relate to the "things readily called good", namely that they stand above these things, and that they are not objects of "any great desire or terror" (Seneca 2015, 303). Put like this, it is a point that the Peripatetics will certainly agree with: after all, greatness of mind is, for Aristotle, the completion of virtue, and the great minded person does stand above such things as wealth and honour, seeing them as having little importance. This is an important point: both the Stoics and the Peripatetics are going to agree that the virtuous person's use of wealth should be reasonable, and that the virtuous person will spend money thoughtfully and accurately. But they also both agree that the virtuous person is impressive and awe-inspiring precisely in attaching very little importance to money. Of course, this does not show that wealth has no value whatsoever, and Peripatetics will insist that it has some small value, and that this value is reflected in the care that the virtuous person takes in its administration. And here, Seneca is not in much of a position to develop a case from the perspective of the virtuous person: as we saw above, the indifferents become good in the hands of a virtuous person; it is enough for us to note here that they become good in great part through being held in disdain.

What comes next are two examples of supposedly bad people who owned a lot of money. Seneca's examples are rather irritatingly chosen: Chelidon, whose only fault seems to be that he was one of Cleopatra's eunuchs, and Natalis, a person whose only fault seems to be that he gave oral sex to women (*EM* 87.16). For the purposes of understanding this exercise, we ought to substitute our own examples of wealthy bad people: wealthy gladiator trainers, the owners of battery farms, cigarette company executives, Donald Trump, or whoever else you might think of as being awfully wealthy. Seneca invites us to make two observations about wealth in these cases.

The first point is contained in the question of whether bad people are defiled by wealth, or whether their wealth is defiled by them (*EM* 87.16). He does not expand on this point much, but rather returns to the comparison with virtue: virtue does not care much for these things (*EM* 87.17). Presumably, one of the things we are to see here is that bad rich people care

too much about their money, so that they appear corrupted by it. In fact, Seneca has a different account of the causal relationship here: it is vice that gives money its corrupting force. Indeed, if we return to the examples given in the syllogism, that of pimps and of gladiator trainers, these are two classes of people who—at least in the ancient world—were willing to engage in profound forms of exploitation in order to increase their wealth. It is probably also important, and tied in with another Stoic argument strategy, that their wealth enables them to carry out their injustice on a greater scale: hiring more gladiators, or in the modern examples, expanding the battery farm, or paying bribes. The Peripatetic will certainly agree that money in these cases becomes both a motive for evil and an enabler of it, and that they would not want to have money *in that way.*

A natural thought to have at this point is that we would still rather like to take the wealthy person's money and distribute it to the poor, or even to have it for ourselves. This is no doubt true, and Seneca would agree with the sentiment. But remember that, for Seneca, indifferents changes valence depending on whether they come into contact with vice or virtue. In the case of donating the money, it would be important for a Stoic not to see this is as a great act of charity, but simply as letting go of something with little value: when a wealthy person donates large quantities of money, the greatness in this is their recognition that they are actually not giving up anything particularly valuable in doing so. (*De Ben.* 12.3--14.3) The money that is dirty in the hands of the gladiator trainer becomes cleansed by its contact with virtue. Similarly, when we imagine having the money for ourselves, we presumably do not imagine having it greedily, or deploying it for some nefarious scheme; or if we do, then we are not the well-intentioned Peripatetics to which this argument is addressed.

Seneca has adroitly used commonalities between the Stoic and the Peripatetic conception of a virtuous person to develop something akin to a thought experiment about the value of wealth. The exercise is to alternatively imagine it in the hands of someone virtuous and in the hands of someone vicious, and to observe how its value appears to change; furthermore, we must observe how lightly the virtuous person holds it. Strikingly, this is precisely the same procedure we found in *EM* 82.12, where Seneca instructs us to compare Cato's case with that of Decimus Brutus. And it is notable that the syllogism forms the blueprint for this exercise: filling out the syllogism requires us to think of bad people who have the good that we are considering. In this sense, the syllogism is not too different from Zeno's syllogism about death; after all, Zeno's syllogism makes for a fair blueprint for a thought experiment as well; namely, to consider glorious instances of a supposedly bad thing. However, in the act of refutation, a dialectician

actively prevents the interlocutor from considering the thought experiment clearly, by leading their attention down the "miniscule and thorny paths" of how to escape refutation and while saving appearances (*EM* 823.22). In the discussion of the syllogism from *EM* 87, Seneca shows us an alternative, and arguably more fruitful, way of finding and using common ground in an argumentative exchange.

Let us turn now to the second syllogism we will consider:

> That which is good does not come of what is bad.
> But wealth comes of avarice.
> Therefore wealth is not good. (*EM* 87.22)

Again, considered from a dialectical perspective, this syllogism is rather useless, which is probably why Inwood (2007) asks, "Why would Seneca advance such a weak argument?" For anyone who believes that wealth, pleasure, or even knowledge is good would reject the major premise as being clearly false. This is a premise that one might expect only an already committed Stoic to accept, and more as a corollary of the theory of the good than as a lemma on the way to its demonstration.

As Inwood (2007) noted, Seneca here seems to shift from his engagement with the Peripatetics to an engagement with the Epicureans. For the objection, he considers temple robbery as being overall bad: certainly, temple robbery produces profit, but it also produces "fear, anxiety, torments of mind and of body" (*EM* 87.22). This looks like a fairly standard Epicurean analysis of the importance of justice: we need justice to live together peacefully, and we should avoid injustice for the consequences it has on our tranquillity, rather than because it is bad in and of itself.

Seneca's discussion of this objection takes two parts. In the first part, he argues, very aptly, that this is simply psychologically implausible. He points out first that many people are unashamed of stealing, and even go so far as to boast of adultery (*EM* 87.23). His strongest point, however, is a political one of some contemporary relevance. He points out that great acts of temple robbery "are carried in a triumphal parade (*EM* 87.23)". This is a strong point. Although private citizens are subject to the laws of the state, and can only very rarely act with complete certainty of avoiding punishment, powerful states often do act in violent and unjust ways without any real fear of reprisal, either against other states or against populations within their own state. If we rely on the psychological consequences of injustice to speak against it, it will be hard to see how to argue against such injustices. Because the psychological consequences of injustices are inconsistent, they cannot be what makes injustice bad, because it is consistently bad. Here Seneca is trying to build common ground through the use of appeals to experience;

what we find is not a refutation of the Epicurean position from premises that they will accept, but observations that make the Epicurean position look less plausible.

Seneca's second point starts out as a fairly standard Stoic *reductio:* if temple robbery is good in some respect, then it will be honourable in some respect, and if honourable in some respect, then right in some respect (*EM* 87.24). This is akin to the argument that Cicero would famously describe as a "leaden dagger" at *De Finibus* IV.18, where he complains that it is dialectically ineffective, as the Stoics' opponents will not accept the step from good to honourable. Again, we might be concerned that Seneca has forgotten the importance of arguing from common ground. However, he proceeds to present a thought experiment: if you were to excuse temple robbery from all punishments and guarantee it safety, then temple robbery would be good, but—surely the Peripatetics and Epicureans agree—temple robbery is never good. Here, the interlocutor is asked to consider a world in which temples can be safely robbed: it would still be wrong to rob the temples. In presenting such a thought experiment, Seneca is not, as Cooper complains, making an appeal to emotions or giving up philosophy in favour of pastoral care: he is searching for some common ground, not with a view to forming a refutation, but simply with a view to finding real common ground with his opponents.

The final move Seneca makes is to consider a different Stoic position that uses an analogy of a snake and a coin in a jar. I think the Stoic position here must be the move that the Peripatetics would make as well—Inwood (2007) describes it as a dialectical concession—because the Peripatetics would certainly accept that temple robbery was bad in and of itself. In the analogy, the jar is the temple robbery, the coin is the profit from the temple robbery, and the snake is the injustice of temple robbery. The coin, in this picture, is incidental and separable from the injustice of robbing the temple, and if one can somehow manage to take the coin without committing the evil, it will still be good. So the bad thing in the temple robbery is not the profit, but rather the injustice; the profit can still be a good thing, even if temple robbing is bad.

Seneca's response is, I think, exactly right, but very brief. He argues that the cases are disanalogous. In the case of the jar the snake and the coin are separable. But in the case of the robbery of a temple, the profit one makes from robbing the temple is not separable from the injustice one commits. The injustice of my robbing the temple is, after all, partly constituted by the fact that I profit, and the community for which the temple is sacred loses. It is simply conceptually impossible that I should take the money without committing an injustice.

Seneca has used the syllogism to develop a series of philosophical investigations. If we are considering whether an apparent good is good, we should consider it when it has come from a bad source; we should consider how it would be if it were taken without any fear of bad consequences, and we should consider whether the apparent good is separable from the evil act. These observations have effectively the same effect as the previous ones: Seneca hopes that we will see that the valence of the item under consideration changes when it comes from an evil source, that we can see that indifferents are sometimes good—when they come from a good source—and sometimes bad, when they come from a bad source. Again, Seneca uses a combination of appeals to experience and thought experiments to establish common ground, rather than seeking common ground via the social pressures present in a dialectical exchange; further, Seneca seeks to provide exercises that will encourage contemplation of value, rather than syllogisms that will produce refutations.

III. Thought Experiments and Argument

I would like to close with a small remark about the relationship between thought experiments and argument in Seneca. For Seneca, refutation, understood as the practice of eliciting spoken commitments and driving an opponent to the desired conclusion, is an epistemically unsound way of arguing as it trades on mismatches between conviction and assertion, and perhaps even between what one thinks and what one believes. Common ground can be better found through shared experience or even through shared imaginative exercises in thought experiments. Seneca was looking to replace what is effectively an adversarial or competitive argumentative practice into a collaborative and inquisitive one. If this were the case, then we would expect converting thought experiments into refutations would hold a heavy philosophical cost.

A very famous example may be seen in a comparison of two statements of Singer's arguments about action on poverty. Singer (1973) presents the argument as a refutation in writing. We must accept that people starving to death is bad. The next premise he wants the reader to accept is that "if it is in our power to prevent something morally bad from happening, without sacrificing anything of comparable moral importance, we ought, morally, to do it" (Singer 1973, 231). (231) Singer claims that this premise is "uncontroversial", and it is certainly a difficult premise, dialectically, to reject. One does not want to appear to be a person who will let evil flourish by standing by. But I doubt Singer's readers are, in general, thoroughly committed to this claim, although they may be prepared to assert or even

think it. So, we might follow Seneca's criticism of Zeno in questioning whether this is really arguing from common ground. In this presentation, the famous thought experiment about saving the drowning child appears as a rhetorical, or perhaps dialectical, flourish to try to encourage acceptance of this premise, but it is surely a long reach from the idea that we must get our clothes muddy to save a drowning child to the principle stated above. Singer's article (1973) is aimed at philosophers, and has spurred an enormous philosophical discussion that is strikingly reminiscent of the discussion of Zeno's syllogisms: it is possible that many of Singer's readers feel trapped by the argument, and feel forced, dialectically, to say that which they do not believe, and are therefore, rather than really thinking about poverty, looking for a way to escape refutation while saving face.

A presentation in *New Internationalist* takes a different approach (Singer, 1997). This article is aimed at a general, rather than a philosophical audience. It does not aim at refuting them, but rather at developing the potency of the thought experiment on the drowning child, and pushing for consideration of what our obligations are to others, given our increasing ability to know what is happening far away and our ability to act at a distance. Being aimed at a public audience, it is not a piece of professional philosophy like Singer's 1973 article. Nevertheless, as philosophy, it has one serious advantage: it argues from ground much more firmly held in common with the reader. After considering Seneca's reflections on dialectic, I am inclined to think the later piece is not only a more rhetorically effective article, but also, contrary to appearances, philosophically sounder.

Bibliography

Barnes, Jonathan. 1997. *Logic and the Imperial Stoa*. Leiden: Brill.

Barney, Rachael. 2003. "A Puzzle in Stoic Ethics". *Oxford Studies in Ancient Philosophy*, Vol. 24: 303-340.

Castagnoli, Luca. 2010. "How Dialectical was Stoic Dialectic?". In *Ancient Models of Mind*, edited by Andrea Nightingale and David Sedley, 153-179. Cambridge: Cambridge University Press.

Cooper, John. 2004. "Moral Theory and Moral Improvement: Seneca". In *Knowledge, Nature, and the Good*, 309-335. Princeton: Princeton University Press.

Duncombe, Matthew and Dutilh Novaes, Catarina. 2016. "Dialectic and Logic in Aristotle and his Tradition". *History and Philosophy of Logic*, Vol. 37, No. 1: 1-8.

Inwood, Brad. 2007. *Selected Philosophical Letters.* Oxford: Oxford University Press.

Kakkuri-Knuutilla. 2005. "The Relevance of Dialectical Skills to Philosophical Inquiry in Aristotle." *Rhizai.* 1: 31-74.

Klein, Jacob. 2015. "Making Sense of Stoic Indifferents." *Oxford Studies in Ancient Philosophy.* 49: 227-81.

Seneca. 2015. *Letters on Ethics*, translated by Margaret Graver and Anthony Long. Chicago: University of Chicago Press.

Singer, Peter. "Famine, Affluence and Morality". *Philosophy and Public Affairs*, Vol. 1. No. 3: 229-243.

Singer, Peter. 1997. "The Drowning Child and the Expanding Circle". *New Internationalist.* Accessed December 10, 2020. https://newint.org/features/1997/04/05/peter-singer-drowning-child-new-internationalist.

Slomkowski, Paul. 1997. *Aristotle's* Topics. Leiden: Brill.

Smith, Robin. 1993. "Aristotle on the Uses of Dialectic". *Synthese*, Vol. 96, No. 3: 335-358.

Smith, Robin. 1997. *Aristotle's Topics I & VIII.* Oxford: Oxford University Press.

Williamson, Timothy. 1994. *Vagueness.* Oxford and New York: Routledge.

CHAPTER SEVEN

HOW TO RECONSTRUCT A THOUGHT EXPERIMENT[1]

MAREK PICHA

Introduction

Although thought experiments are widely used in the sciences, in philosophical arguments as well as in everyday communication, there is no consensus among scholars about their nature. It is striking that such a common method has no generally accepted definition and no set of sufficient and necessary characteristics. It is not clear what a thought experiment is and how to distinguish one from other forms of speculation. One cannot rely on a widely recognized typology of thought experiments. There is no consensus about the way a thought experiment reaches its goal, and it is not even clear what the goal is supposed to be. It is difficult to recognize a thought experiment. It is even more difficult to theorize or come to an agreement about one. Different authors emphasize different aspects: some emphasize the similarity of design between thought experiments and real experiments; others point out their relevance to cognitive models, works of art and mnemonic devices.[2] I will treat thought experiments as narrative structures that use scenarios with particular details and hypothetical

[1] This paper has been written with the aid of the Czech Science Foundation, grant No. 401/09/0661, and was originally published in *Organon F*, Vol. 18, No. 2: 154-188, 2011. Also, I would like to thank Dagmar Pichová and Radim Bělohrad for their help and their thoughtful comments.

[2] The defence of thought experiments considered as an evolutionary stage of real experiments can be found especially in Sorensen (1992). For the conception of thought experiments as mental models see Nersessian (1993). The relation between literary fiction and philosophical thought experiments is analysed for example by Camp (2009). Ernst Mach studies the role of thought experiments in the process of recollection and explication; see Mach (1905).

premises to give an answer to a presented question, but where none of the scenarios must be realized to reach the goal.

This text is a contribution to the debate on the epistemological status of thought experiments. How can hypothetical examples participate in the process of obtaining knowledge? Is it ever possible to accept an imaginary scenario as a source of justification, that is to say, to accept it as a good reason to embrace an opinion? These questions have been a source of lively debate[3] and have resulted in a division of scholars into those who think that thought experiments are epistemologically important concepts and those who consider them uninteresting.

The difference in their opinion of the epistemological status of thought experiments can be demonstrated by two questions: 1.) Are thought experiments sources of *new* knowledge? 2.) Are thought experiments *unique* sources of knowledge? The first question concerns if thought experiments can justify beliefs that cannot be justified by other means prior to the experiment. Is drawing conclusions from imaginary scenarios merely reformulating what is already known? Are thought experiments merely aids facilitating the acceptance of views we have already accepted and observed, but refused or failed to accept consciously? Although I will touch upon these issues at several points, the second question will be my primary concern.

Epistemological uniqueness can be understood in various ways. Are thought experiments structures enabling us to obtain information that we could not obtain by other means? Are they structures that make use of our otherwise latent cognitive functions? Is a thought experiment basically an idiosyncratic and irreducible structure? None of these questions will be the topic of this text. What I am interested in is whether the epistemological importance of thought experiments can be identified with the epistemological importance of other, less elusive structures. I am interested in the epistemological uniqueness of experiments in the sense of their irreducibility to other sources of justification—I will, in particular, criticize an influential argument for the irreducibility of thought experiments to arguments. First, I will introduce the radical empiricist theory of eliminativism, which considers thought experiments to be rhetorically modified arguments, uninteresting from the epistemological point of view. Then, I will present objections to the theory, focusing on the critique of eliminativism by Tamar Szabó Gendler, analysing her objections and showing their drawbacks.

[3] Brown (1991a), Norton (1996), Gendler (1998), Bishop (1999), Borsboom et al. (2002), Häggqvist (2009). Only a partial list is presented here. The discussion is very rich.

Empiricism and Thought Experiments

Thought experiments pose a challenge for empiricist epistemology, i.e. for the theory that all true synthetic beliefs must be directly or indirectly grounded in sensory experience only. Leaving aside analytic truths, the only ultimate source of epistemic justification is, for an empiricist, sensory experience. Thought experimenting is, obviously, incompatible with this project, since it is a way of obtaining information from the armchair, i.e. lacking the relevant perceptions that could justify our beliefs. There is no doubt that there have been plenty of cases in the history of science and philosophy in which the evaluation of a belief or a theory was based on the supposition of fictional events, the mental manipulation with imaginary objects, and the investigation of hypothetical states of the world. Thus, thought experiments have traditionally been conceived as good reasons for accepting or rejecting a standpoint, which deserves critical attention from the empiricist: If thought experiments provide us with knowledge of our world, where does this knowledge come from? How can we explain the epistemological status of thought experiments as well as remain loyal to the tenets of empiricism?

Of course, the empiricist cannot admit that thought experiments lead to knowledge of contingent things by means of rational inquiry. Such sources of justification are not acceptable; knowledge must be derived from sensory perception. Ernst Mach was an initiator of the attempts to expose the epistemic value of thought experiments in the framework of empiricism. In his *Science of Mechanics*, he presents a conception according to which thought experiments are tools that enable us to bring to surface our hidden beliefs (Mach 1960, 27-28). Mach supposes that not all of the information obtained through sensory perception is used to form explicit beliefs, much of it is processed on the unconscious level for which Mach uses the term "instinctive". Our minds contain imaginary stocks with the well-lit areas filled with reflected, explicitly embraced beliefs. Besides those, however, there are dark corners, whose contents are unknown but which influence our behaviour and decisions. Thought experiments are one way of bringing beliefs from the dark corners into the light, i.e. they enable us to turn instinctive knowledge into explicit knowledge.

Mach's model is fully compatible with the empiricist tenets. The role of thought experiments is limited to processing information obtained through the senses. The imaginary scenarios help structure, conceptualize,

and explicate the information.[4] His conception of thought experimenting could be very loosely associated with hypnosis, in which the subject is able to recollect details of past events of whose existence she has been completely unaware. Mach naturalizes the experiments. His solution to the problem is based on the understanding of thought experiments as ways of processing information coming from unquestionable sources: experiments provide us with new knowledge of the world as they are intermediaries between perception and consciousness. Thought experiments without prior perception are, in fact, no thought experiments at all, they are different modes of thinking or utilizing imagination.

What are the answers that Mach's naturalist explanation strategy can give to the two epistemological questions raised above? With regard to the first question, i.e. whether thought experiments are sources of new knowledge, naturalism distinguishes between an externalist and an internalist variety. These varieties differ in their opinions on the requirement for the accessibility of the justification process to the subject. In other words, the question is if the subject must be aware of having an instance of knowledge. Externalists claim that the subject has a given knowledge prior to the experiment—even if in a tacit form, nevertheless influencing her decision-making and behaviour. According to externalist empiricism, thought experiments are not sources of new knowledge. They only change some properties of old knowledge. Internalists, on the contrary, consider as knowledge only those beliefs that are justified by a process cognitively accessible to the subject. If the subject is not aware of the reasons that justify her belief, her belief is not knowledge. It is impossible to know without knowing that one knows. Internalist empiricists consider thought experiments as genuine sources of knowledge, since it is only the experiments that bring the subject to the acceptance of the justified true belief.

The proponents of empiricism also differ in their answers to the second question, i.e. whether thought experiments are unique sources of knowledge. On the one hand, we find enthusiasts, who are convinced of an irreducible epistemological importance of thought experiments. Thinking over imaginary scenarios cannot, in their opinion, be fully replaced by another source of justification. Thought experiments offer a unique method of obtaining knowledge. On the other hand, there are eliminativists, who consider thought experiments uninteresting from the perspective of the theory of knowledge and, in a sense, epistemically parasitic.

[4] Mach distinguishes between implicit beliefs obtained by means of personal sensory experience and innate implicit beliefs – those are, however, ultimately obtained by means of our ancestors' (personal) sensory experience.

1.1 The eliminativist thesis

According to the eliminativist version of empiricism, thought experiments have no unique and independent epistemic power and are, in fact, uninteresting as a method of obtaining and justifying beliefs. Eliminativists consider thought experiments to be epistemologically marginal extensions of other, less problematic sources of justification. In particular, thought experiments are just dressed-up arguments and what is interesting about them with respect to the justification of beliefs can be fully derived from their argumentative core. The most prominent proponent of eliminativism is John Norton, who describes the relation between an experiment and an argument in the following thesis:

> Thought experiments can be reconstructed as arguments based on hidden or explicit assumptions. The resulting belief can be considered justified only to the extent that the reconstructed argument is capable of justifying its conclusion. (Norton 2004b, 1142)

It is a radical opinion, according to which the justificatory power of thought experiments is no stronger than that of the corresponding arguments stripped of the particularities of their experimental design. That does not mean thought experiments have no epistemic power whatsoever. It is only that they have no special and unique role. If we strip a thought experiment of its particular sets, actors, and attractive plot, its justificatory power does not change. Of course, we lose what is attractive about thought experiments, but the particular details of the experiments do not add any epistemologically relevant features to the arguments.

Tamar Gendler, whose critique of eliminativism I will focus on, refines the key thesis (Gendler 1998, 398 and further). First, she points out two possible but incorrect interpretations of the eliminativist thesis. The first interpretation is that a straightforward argument, that is, an argument without particular premises, can be used to derive the same conclusion as can be derived from an experimental scenario. This interpretation makes the eliminativist thesis trivially true and every physics textbook proves that. There is no doubt that the conclusion we derive from a thought experiment can be derived from a straightforward argument as well. The other incorrect interpretation of the eliminativist thesis is that a person who understands the conclusion of a thought experiment could have demonstrated to her the same conclusion by means of an argument. This interpretation is trivially incorrect as it ignores the extraordinary didactic qualities of thought experiments. A thought experiment may reveal to a person what a straightforward argument may not be able to. The first incorrect interpretation is too weak; the other is

too strong. To prevent possible misunderstandings, Gendler formulates the eliminativist thesis in the following way: reasoning about specific entities within the context of an imaginary scenario does not lead to a rationally justified conclusion that would not be rationally justifiable on the basis of a straightforward argument based on the same initial information.

Thus, the issue is whether particular details influence the process of belief justification. Eliminativism claims that if we can talk about justification in the context of thought experiments, it is the underlying straightforward argument that does the job. A straightforward argument is able to justify the conclusion with the same strength as the thought experiment if it is based on the same premises. If we have the same initial conditions, the absence of particular details has no impact on the justificatory power. Enthusiasts, on the contrary, say that the loss of particular details leads to a loss of justificatory power.

The dispute between eliminativists and enthusiasts takes place in the context of Galileo's famous thought experiment with falling objects. The bone of contention is the sufficiency of the argumentative reconstruction of this imaginary scenario launched against Aristotelian physics—Galileo's example is put forth as a model of a great, cogent thought experiment in which the loss of particular details would lead to the loss of epistemic power. I will present James Brown's reconstruction of Galileo's experiment in Chapter 2, and then the way John Norton replies to the critique. Chapter 3 deals with a sophisticated critique of eliminativism by Tamar Szabó Gendler, who revises Brown's reconstruction of Galileo's experiment, supplements it with potential opponents' replies, and shows how the robustness of the thought experiment differs from the robustness of its straightforward reconstruction. She then generalizes the identified difference and puts it forth as an argument against eliminativism. In Chapter 4, I will first reconstruct Galileo's experiments using argument diagrams and then a simplified Toulmin model of an argument. I will show that Gendler's critique of eliminativism is based on a simplification and an inadequate description of the relevant characteristics of Galileo's experiment. I will show a way of defending eliminativism from the charge based on this particular thought experiment.

Brown's Critique of Eliminativism

James Brown's rationalist conception is an alternative to the eliminativist attitude. Brown believes thought experiments are tools that enable direct access to the ideal world of physical laws. He answers the question of how purely rational activity can lead to the acquisition of new empirical

knowledge by postulating the existence of a special epistemic channel between reason and the system of independent abstract entities whose relations constitute the laws of nature. His views are in sharp contrast to empiricist epistemology and his Platonism is, thus, a parallel rival theory to eliminativism. The main topic of this paper is, however, not the polemic between empiricism and rationalism, but the polemic between eliminativism and the rest of the world, the polemic to which Brown has also contributed in a way that is not seriously contaminated by his unorthodox epistemological views. Brown is considered to be an influential scholar for his advanced typology of thought experiments, among other things, which he introduced in his monograph (Brown 1991a, Chapter 2). This typology is a suitable starting point for the introduction of his critique of eliminativism, as he uses it to identify those experiments that resist elimination. He believes there is a group of experiments whose epistemic value cannot be fully represented by a set of general premises and a conclusion.

Brown distinguishes destructive and constructive experiments. Destructive experiments provide counterexamples to a theory; constructive experiments are meant to support one. The support can have three forms. Firstly, we have a theory and the experiment is an example that can illuminate the theory and help apply it to particular phenomena. The experiment plays the role of an illustration. The situations and phenomena used in the experiment are unproblematic and usually refer to ordinary experience. Brown calls these experiments mediative. Secondly, we do not have a theory; we are looking for one. The thought experiment presents an unusual or speculative phenomenon that we attempt to explain. This type is called conjectural. Thirdly, Brown speaks about direct thought experiments, which result from a combination of both of the abovementioned types. They share the lack of a theory with conjectural experiments and the unproblematic character of the phenomena we are trying to explain with the mediative experiments. Direct thought experiments draw attention to a phenomenon that can neither be doubted nor adequately explained. That results in the creation of a suitable explanatory framework.

The identification of conjectural and direct thought experiments is the key point of Brown's critique of eliminativism. Brown believes one cannot form their adequate argumentative reconstruction:

> We have clearly specified premises to work from in either destructive or mediative examples; but in the case of either direct or conjectural thought experiments we simply do not have a definite background theory from which we can be said to be arguing to our conclusion. (Brown 1991a, 47)

Some thought experiments are not based on a well-formed theory that forms a derivative basis of the argument. Brown assumes that argumentative reconstruction must be in the form of a derivation of a conclusion from premises that, among other things, contain the hypothesis that serves as an explanatory framework of the phenomenon established in the experiment. I confess that I do not clearly see what justifies this assumption. Brown believes that the only adequate structure of a reconstructed argument is the following: Considering phenomenon P under theory T, conclusion C follows. I believe this conception of argumentative reconstruction, i.e. the conception of what kind of argument the reconstruction should be, is too narrow. It seems that Brown means by reconstruction: (a) the formulation of a deductive argument, where (b) all premises must already be explicitly formulated in the unreconstructed form. It follows, then, that experiments in which the theory is derived inductively or abductively cannot be reconstructed (ad a). Further, experiments in which the theory is not explicitly introduced among the premises cannot be reconstructed (ad b). Brown's interpretation of the eliminativist thesis is not in accordance with its intended sense. The conditions that Brown states are not part of it; Norton explicitly denies them. Eliminativists do not assert that all thought experiments can be reconstructed as deductive arguments without a loss of epistemic power, but they do claim that all thought experiments can be reconstructed as deductive or inductive arguments whose epistemic power does not change if particular premises are removed from them. Neither is it asserted that the premises of the reconstruction should only include the explicit statements given in the experiment. The eliminativist thesis explicitly speaks about "hidden" premises.

Brown's demand for the derivation of the conclusion from a well-formed theory is too strong, since the theory does not always need to be contained in the premises. His structural objection is based on an inadequate understanding of what is meant by the argumentative reconstruction of a thought experiment. The refutation of the objection is not particularly difficult. It is sufficient to point out the misinterpretation of the criticized view. However, this is not the only objection Brown raises against eliminativism. Another of his objections concerns the ability of the reconstruction to represent all of the epistemological contribution of the experiment. Brown asks whether the reconstruction by means of a straightforward argument leaves something important out. His typology of thought experiments is relevant here again. The last item in his typology are experiments that he calls Platonic. They are experiments that fall into two of the abovementioned categories. They are both destructive experiments, as their role is to reject a theory; and direct experiments, which establish a

new theory by means of an unproblematic phenomenon. A Platonic thought experiment is a scenario in which thinking over hypothetical but relatively common situations leads to the disclosure of drawbacks in the current explanation and the formulation of a new, better and more adequate explanation. It is supposed to be the highest form of a thought experiment, as it shares the qualities of all the other types: it refutes the old conception and establishes a new one by means of an unproblematic phenomenon. Platonic thought experiments are epistemically richer than straightforward arguments; their contribution cannot be fully represented by a sequence of premises and a conclusion.

2.1 Galileo's experiment with falling objects

Brown claims that it is not possible to reconstruct a Platonic thought experiment by means of a straightforward argument without a loss of epistemic power. He presents the EPR paradox, Leibniz's experiment to prove *vis viva* and Galileo's example of falling bodies as examples of such scenarios. For its ingenious simplicity, cogency and clarity, Galileo's thought experiment has become the focal point of the debate on the epistemic power of thought experiments. Brown reconstructs the scenario and shows where the experiment and the argument, in his opinion, gap.

Galileo's experiment attacks Aristotelian physics, and in particular the view that the natural speed of bodies is directly proportional to their weight. Aristotle claims that heavier bodies will fall more rapidly than lighter bodies and Galileo's literary projection raises doubts about whether Aristotle actually verified his statement empirically. It would surely be possible to conduct an experiment to confirm the truth of the claim, but it is not really necessary. The falsity of the Aristotelian principle can be shown without a real experiment, says Galileo. It is sufficient to reason as follows: let us assume, together with Aristotle, that bodies of different weights fall at different speeds in the same medium—if we take two bodies of different weights, the heavier body will fall more rapidly than the lighter one—at what speed will the connection of the two bodies fall?

> Then if we had two moveables whose natural speeds were unequal, it is evident that were we to connect the slower to the faster, the latter would be partly retarded by the slower, and this would be partly speeded up by the faster. /.../ But if this is so, and if it is also true that a large stone is moved with eight degrees of speed, for example, and a smaller one with four [degrees], than joining both together, their composite will be moved with a speed less than eight degrees. But the two stones joined together make a larger stone than the first one which was moved with eight degrees of speed;

> therefore this greater stone is moved less swiftly than the lesser one. But this is contrary to your assumption. So you see how, from the supposition that the heavier body is moved more swiftly than the less heavy, I conclude that the heavier move less swiftly. (Galileo 1974, 65)

Galileo constructs a *reductio ad absurdum* in the experiment. He assumes Aristotelian dependence of the speed of a body on its weight and shows that the assumption leads to unacceptable results. One material system would have to fall at two different speeds. How can this thought experiment be transformed into an argument? Brown suggests the following reconstruction:[5]

[Argument A: Brown's reconstruction]

1. The natural speed of falling bodies in a given medium is proportional to their weight.
2. If a large stone is moved with eight degrees of speed, a smaller stone with half the weight will be moved with four degrees of speed.
3. If the slower stone is connected to the faster stone, the slower one will decelerate the faster one and the faster one will accelerate the slower one.
4. If the two stones from premise 2 are connected, the resulting object will fall more slowly than at eight degrees of speed.
5. The weight of two connected objects is higher than the weight of the bigger of the two objects.
6. The connection of the stones from premise 2 will fall faster than at eight degrees of speed.
7. Premises 4 and 6 contradict each other.
8. Thus, premise 1 must be rejected.
9. Thus, all stones fall alike.

Leaving aside the fact that the proposed reconstruction contains particular premises,[6] this argument represents precisely what is attractive about Galileo's example from Brown's perspective. The reconstruction shows that Galileo's thought experiment has an unproblematic design, since with the exception of the initial Aristotelian premise necessary for the reductio, there are no controversial statements. It is a Platonic thought experiment: statements 1-8 describe its destructive component; statement 9 is a constructive step leading to the establishment of a better theory. It is this very step from the rejection of the Aristotelian thesis to the acceptance of the Galilean thesis that, according to Brown, presents an insurmountable

[5] According to Norton (1996, 341 ff.).
[6] Namely premises (2), (4), and (6), whose dispensability will be dealt with below.

challenge for the eliminativist conception. What is the challenge? The move from 8 to 9 is neither an inference nor an inductive generalization grounded empirically. Nevertheless, after careful consideration of the experiment, this move is believed to be justified and hardly anyone would hesitate to make it. Brown sometimes refers to this move as a "Platonic leap" that cannot be represented by a straightforward argument as a legitimate move from the premises to the conclusion. Eliminativists are bound to regard this move illegitimate as it is not sufficiently supported by the premises of the reconstruction. It is, however, completely acceptable in the context of the thought experiment. The epistemic value of the experiment parts with the value of its reconstructed form at this point—the experiment is richer.

Norton answers this critique in two steps. In the first step, he supplements Browns reconstruction with implicit elements that serve to represent Brown's interpretation of the experiment more precisely. In the second step, he shows that (i) these elements enable us to analyse the Platonic leap as a straightforward argument, and that (ii) Brown's interpretation of Galileo's experiment is incorrect.

As I have said, the Platonic leap is supposed to occur between the destructive premise/intermediate conclusion 8 and the constructive conclusion 9. A common interpretation of Galileo's experiment, one that, according to Norton, Brown would share, works with a hidden assumption that to determine natural speed, it is not necessary, according to Aristotelian physics, to consider any quantities other than the weights of the falling bodies. In other words, natural speed depends *solely* on the weight of the falling bodies. Norton believes that if we put this hidden assumption into the reconstruction, no Platonic leap is needed and the conclusion can be reached by a simple inference. The key step from 8 to 9 can be reconstructed as a straightforward argument:

[Argument B: Norton's reconstruction of the Platonic leap]

(8a) The natural speed of falling bodies depends only on their weight.

(8b) The natural speed of falling bodies is some arbitrary, monotonously rising function of their weight.

(8c) If the function is anywhere strictly increasing, then we can find a composite body whose natural falling speed is intermediate between the falling speed of its lighter components.

(8d) Premise 8c is incompatible with premise 8b.

(9) Thus, the function is constant and all stones fall alike.

Norton presents the Platonic leap as an ordinary inference from implicit and explicit premises. 8a contains an implicit assumption about a strict dependence of natural speed on body weight. 8b is a specification of the dependence. It says that whenever the weight increases, the speed must increase as well. It further says that, for the purposes of the inference, it does not matter at what rate or according to what factor the quantities increase. What is important is that the weight of an object cannot increase while its speed remains constant. Premise 8c is deduced from 3, 5 and 8b. A compound body falls, according to the third premise, more slowly than its parts. According to the fifth premise, the weight of a compound body is always higher than the weight of any of its parts. Premise 8b says that there can be no exceptions with very heavy or very light parts. The intermediate conclusion 8d claims the incompatibility of premise 8b and the inferred conclusion 8c: if speed must accelerate with any increase in weight, the same must hold for the increase in weight when two objects of different weights are connected. According to 8c, however, the speed will not increase in such cases. Conclusion 9 says that premise 8b is false, as it leads to a contradiction. This way one can derive a constructive conclusion from Galileo's experiment. The natural speed of a body is independent of its weight, because to suppose otherwise leads to a contradiction. The new hypothesis is not formulated by means of a mysterious insight into the world of the laws of nature. It is a plain inference from premises. If we supplement the reconstruction of the argument with implicit premises 8a and 8b, we can show that the conclusion is not a Platonic leap, but a simple inferential step. Norton points out another interesting thing: premise 8a is not even implicitly contained in Galileo's experiment. The above stated enrichment of the reconstruction does represent a way of deriving a new theory from the experiment, but this derivation is not sufficiently grounded in the experiment itself. Where is the problem? Premise 8a expresses an idealized situation of natural speed in vacuum. In such a situation, one need not consider differences in speed caused by different aerodynamic shapes of falling bodies, and natural speed is, given Aristotelian principles, solely a function of weight. Galileo's experiment is, however, explicitly designed for bodies falling in a medium—as stated in premise 1—and it is illegitimate to take the step from 8 to 9 under such circumstance. The interpretation of the experiment described above is, thus, an anachronism. We are only confronted with the scenario ex post, we know its purpose and know the relevant polemic to an extent. Galileo's goal in the experiment was merely to rebut the Aristotelian conception of natural speed of bodies in a medium, not to create his own theory. His experiment is not sufficient for that purpose. What is, then, Norton's reply to Brown's objection? Galileo's

example is not an instance of a Platonic experiment! The step from 8 to 9 is not sufficiently justified by the experiment; we only accept it, because we have been instructed to do so. And we do that by inferring the desired conclusion from some implicit premises.

Gendler's Critique of Eliminativism

Tamar Szabó Gendler has formulated a noteworthy critical reaction to eliminativism.[7] Like Brown, she attempts to show that thought experiments are epistemically richer than straightforward arguments. She also makes use of Galileo's example of falling objects to demonstrate the inadequacy of the argumentative reconstruction of the experiment. She bases her critique of eliminativism on the same view: while an inferential step is epistemically justified in a thought experiment, it is illegitimate in the argumentative reconstruction of the experiment. Brown defends the uniqueness of Galileo's experiment by pointing out that it leads to the *formulation of a better theory*. Gendler bases her defence on the idea that, unlike the argument, the experiment can tell us *what is wrong with the original theory*.

The basis of her critique of eliminativism is a minimalist reconstruction of Galileo's experiment. She removes all premises that are, in her view, irrelevant, and puts the remaining ones into the following argument:

[Argument C: Gendler's reconstruction]
(I) Natural speed is mediative.
(II) Weight is additive.
(III) Thus, natural speed is not directly proportional to weight.

Statement I is a general principle concerning the interaction of two connected falling bodies with different weights. The corresponding premise in Brown's reconstruction is premise 3. Statement II is a reformulation of premise 5 of Brown's initial reconstruction. Statement III contains the conclusion of the argument, to which the intermediate conclusion 8 above corresponds. Particular premises 2, 4 and 6 have been omitted; the remaining ones have been formulated as generally as possible.

Gendler believes that conclusion III receives different degrees of justification from the thought experiment and from the corresponding straightforward argument. The key idea is that while the straightforward argument offers a number of ways to improve the refuted Aristotelian theory, the thought experiment reduces them to a single one. The thought experiment can tell us that something is wrong with the original theory, as

[7] Gendler 1998.

well as reveal the problematic point. The straightforward argument is less potent in this respect**.** We can look for the problematic aspect of the refuted theory basically anywhere in the context of the argument. The crucial difference between Gendler's and Brown's critiques of eliminativism lies in a different target of their objections. Brown disputed the move from intermediate conclusion 8 to conclusion 9*;* Gendler attacks the move from premise 7 to intermediate conclusion 8. Let us look at her critique in a greater detail.

Assume you are a proponent of Aristotelian physics, who would like to reply to Galileo's example and revise your theory of natural speed. There are several ways to supplement or refine the theory, so that it could avoid the paradox described in the experiment. Gendler identifies what she calls four ways out, that is, four procedures the proponent of Aristotelian physics can employ to save the hypothesis that natural speed is proportional to weight.

(C1) Natural speed is not physically determined for connected bodies.

(C2) Weight is not physically determined for connected bodies.

(C3) Natural speed and weight are mediative for those connected bodies that are *united*. Natural speed and weight are additive for those connected bodies that are *unified*.

(C4) The natural speed and weight of connected bodies are determined by the rate of their connection.

Statement C1 introduces an exception in the calculation of natural speed of free falling bodies. The dependence of speed on weight only concerns individual bodies. If two bodies are connected, their resulting speed is not determined by the relation. The scope of the given physical theory is limited here—the Aristotelian theory is valid, but only for unconnected bodies. When they are connected, their speed is no longer a function of their weight. The same idea is used in C2, but the relation in question changes there not due to a change in the speed of the fall after the connection, but because the connection affects the calculation of the total weight. Both cases offer the Aristotelian a solution to the paradox. The key idea of Gendler's critique is that while the proposed solutions are acceptable in the context of the straightforward argument, they are illegitimate in the context of the thought experiment. When we conduct the thought experiment, we simply ignore the ways out. When we evaluate the straightforward argument, we cannot reject the exception in the same way.

The same critical procedure is used in the remaining two ways out. C3 introduces a new distinction in the category of connected bodies. Unified bodies are such that they retain their identity after the connection, that is, there are still two distinct bodies. The process of uniting bodies is, however, a process resulting in a single object. The bodies in Galileo's experiment are merely unified, which affects the calculation of the resulting weight; it is mediative in that case. C4 assumes that the rate of connection of the falling bodies is a relevant physical property and that it determines the dependence of their speed on their weight.

Some of the proposed ways out are relatively sophisticated. Others are based on a robust change in the view of how nature operates. The particular character of the ways out is, however, not essential for the critique of eliminativism. The point of the presented ways out is to demonstrate the idea that the Aristotelian has, in principle, a number of ways to save his theory. While the thought experiment reduces the number beforehand, the straightforward argument does not.

3.1 How to block the ways out?

According to Gendler, the above stated ways out are rejected prima facie when conducting the thought experiment. It is possible to block the ways out with some effort even in the argumentative reconstruction. Additional premises are required, though. The first two ways can be blocked by adding premise:

(D1) Natural speed and weight are physically determined.

This premise states a fundamental principle that the weight and natural speed of connected bodies are quantities fully determined within the physical domain. The formula excludes any extra-physical factors. The connection of simple bodies does not affect the calculation of the resulting weight or speed. The third and fourth ways out are blocked by an additional premise stating what is not part of the physical domain:

(D2) Entification is not physically determined.

That is, the division of a system into individual objects is not one of the fundamental principles of physics. It is up to us how many objects we identify in a given system; it is a matter of a decision whether we consider a system to be a single body or several distinct bodies. The number of objects, bodies, or things is not firmly determined by any physical properties. It is a matter of our point of view. The distinction between

unified and united bodies in C3 is then arbitrary and has no effect on the physical theory. Neither is the rate of connection, the key term in C4, a physically determined property. It is merely a matter of how we describe the system.

The straightforward argumentative reconstruction can, thus, close some ways out of the paradox of the Aristotelian conception too, and manoeuvre the Aristotelian towards a better theory. The point is that what the imaginary scenario achieves without effort the reconstruction can only do by means of complicated metaphysical principles that represent our ideas about the operation of the physical world. According to Gendler, the additional premises are not a matter of course. Their explication in the argumentative reconstruction is a process different from the process of their acceptance in the thought experiment. Gendler says that part of the contemplation of the experiment is a survey of the possibilities for solving the paradox, and some of the possibilities are already rejected in the course of the experiment. The reconstruction achieves the same goal in a more complex way and with additional and controversial premises only. Gendler's argumentation can be summarized as follows:

[Argument D: Gendler's argument against eliminativism]

I) Some objections are refuted when conducting the thought experiment.
II) The same result can only be achieved in the argument by adding premises.
III) Thus, the argument and the experiment differ in the way they refute objections.
IV) Thus, there is an epistemological difference between the argument and the experiment.

Reconstruction: A Diagram

In the following sections, I will defend the eliminativist position against the above objection. Norton's own response is relatively brief. Argument must be able to provide the same results as thought experiment. Otherwise, it would be impossible to consider a thought experiment as a reliable cognitive procedure. Norton creates a dilemma for enthusiasts: either argument is as potent as experiment, which means that a thought experiment might be epistemologically reliable—or experiment provides something more, which however goes beyond its epistemological reliability. Either a thought experiment is a source of new knowledge and therefore it is not epistemologically unique, or it is unique and therefore cannot be source of new knowledge.

Particular details of the scenario may enable us to obtain beliefs that we could not obtain by means of the argumentative reconstruction, or that we could only obtain indirectly. However, that does not mean that beliefs obtained that way are also justified by that process. While Norton's strategy relies on a reliabilist condition of epistemic justification at one horn of the dilemma, I will attempt to present a different view. I will face the objection directly, that is, I will question the view that thought experiments are unique sources of knowledge. I will also show that the conclusion Gendler reaches in her work results from the wrong argumentative reconstruction of Galileo's experiment. The core of Gender's critique of eliminativism is the objection that the recipient obtains a belief in a thought experiment that she may not obtain in a straightforward argumentative reconstruction. There is no doubt about that, since the information presented in the form of a thought experiment is easier to grasp than when in the form of a straightforward argument. Experiments no doubt make obtaining new information easier, and their didactic value is beyond dispute. We commonly and successfully use thought experiments in this way. The question is whether the obvious difference in reception is only caused by individual intellectual abilities of the audience, or whether a contributing factor is that there is an epistemic difference between thought experiments and their argumentative reconstructions. Gendler attempts to show that beliefs obtained by means of a thought experiment cannot be obtained by means of a straightforward argument with the same initial conditions; particularly, some ways out are unacceptable. Using stronger analytic tools, I will show that the ways out are as blocked in the argumentative reconstruction of Galileo's example as they are in the thought experiment. The difference in reception is not caused by an epistemic difference, but a rhetorical one—we tend to overlook some aspects simply because our attention is focused on other aspects.

My defence of eliminativism against the "Galilean" attack is based on the rejection of premise ii in argument D. I claim that the straightforward argument reaches the same goal even without additional premises. The ways out are blocked in the argument by the same procedure as in the thought experiment and no controversial, general, or questionably justified metaphysical premises are needed. I will proceed as follows: I will present and comment on some models of argumentative reconstructions for Galileo's experiment in this chapter. First, I will present a diagram of Brown's reconstruction as a starting point of the polemic. Then, I will present a concise model of Gendler's. Finally, I will propose my own reconstruction, in which I will employ some basic concepts of the Toulmin model to show that the straightforward argumentative reconstruction can block the ways out. The polemic between eliminativists and their critics

over Galileo's example proceeds by means of an unstructured list of premises and conclusions. To highlight the relationship between the elements, I will first write the reconstructions in the form of diagrams. Diagram 1 shows the relationship of the statements in argument A, i.e. the Brown reconstruction that opened the debate.

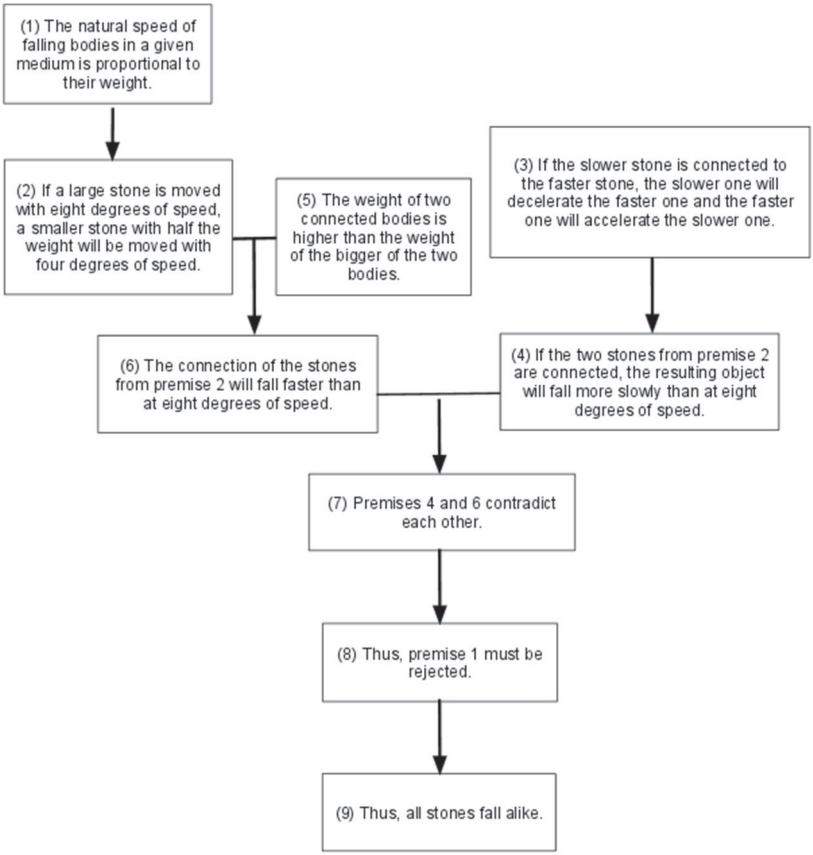

Figure 7-1: Brown's reconstruction

Diagram 1 takes the statements from argument A. They are arranged in an inferential tree. The reconstruction contains the problematic conclusion 9, discussed in detail above. The major problem concerns particular premises 2, 4, and 6. The subject of controversy between eliminativists and the rest of the world is the particular details in the scenario of the thought

experiment. Eliminativists claim that, without the details, the experiment has the same epistemic power as with them. Such a claim cannot obviously be tested if we compare an experiment with particular premises with an argument with particular premises. We have seen that Brown formulated his reconstruction for the purposes of a different critical strategy, and that is probably the reason why it does not meet our demands. For our purposes it must be rid of the particular premises, which I will attempt to do in my own reconstruction of Galileo's example below.

Let us now get back to the alternative model presented by Gendler. Her version of the straightforward reconstruction is put in argument C. This minimalist version is trivially represented by Diagram 2. Gendler formulates premises 3 and 5 so as to emphasize their general character, and skips premises 2, 4, and 6 dealing with particular characteristics of falling bodies.

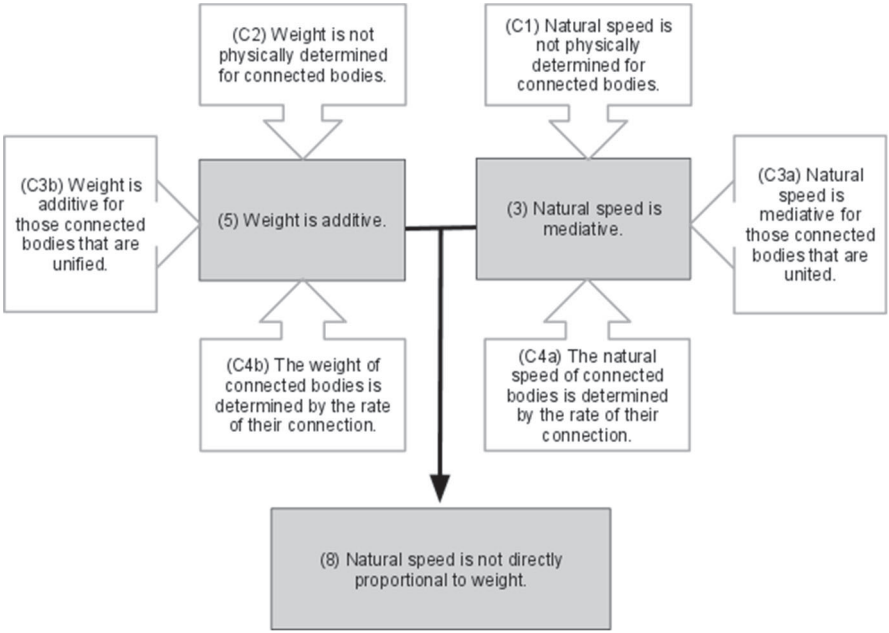

Figure 7-1: Gendler's reconstruction with the ways out

The brief reconstruction is supplemented in the diagram by the ways out that Gendler opens for the Aristotelian advocate. They are types of objections that question the plausibility of the premises. The first way out, C1, attacks premise 3, claiming that natural speed is not physically

determined for connected bodies. It does not follow, then, that the total weight of the compound object will be higher than the weight of the heavier body taken independently. The second way out, C2, states that the weight of connected bodies is not physically determined and attacks premise 5. While the first two ways out speak about speed and weight independently, i.e. each quantity corresponds to one way out, the third and fourth way describe them simultaneously. The reason for this move is not clear, as both ways can further be broken down into two independent objections, i.e. they can apply to weight and speed independently. The third way out, then, attacks premise 5 in version C3a, and premise 3 in version C3b. This yields six ways out in total. I have said that Gendler also presents two theses, two "approximate articulations of defeasible assumptions about the physical world" (Gendler 1998, 408), which block the ways out when added to the reconstruction: statement D1 asserts the physical determination of speed and weight in connected bodies, blocking C1 and C2. Statement D2 then claims the arbitrary character of the connection when calculating the resulting speed, thus blocking ways C3a, C3b, C4a, C4b.

I do not think the proposed reconstruction is suitable for two reasons. First, Gendler does not consider the Aristotelian premise 1. The plausibility of Galileo's example is, besides the unproblematic requisites, mainly based on its form: it is a reductio ad absurdum, i.e. a mode of argumentation in which the premises lead to a contradiction. A strong point from Galileo's example is that the initial premise is the Aristotelian thesis, i.e. the opponent's premise, whose plausibility need not be defended against the Aristotelian. The opponent cannot question its plausibility, which prevents a potential dispute about the matters of fact. Gendler instead reconstructs the thought experiment as an argument based on mathematical reasoning about the relation of two functions with different graphs. The conclusion of the argument is then claimed to be inconsistent with the Aristotelian principle.

Second, the formulation of the premises 3 and 5 is deliberately too brief. The concise statements and terms used give the impression of complicated, controversial claims whose intuitive plausibility can be easily refuted by the ways out. The implausibility of the ways out is based on the acceptability and cogency of the targeted premises 3 and 5. Gendler chooses formulations that decrease the intuitive plausibility. I believe Gendler has thrown the baby out with the bathwater in her generalization. The effort to create a contrast between the particular information in the experiment and the generalizations in the argument results in the fact that her reconstruction fails to contain all the elements needed to assess the argument. The

possibilities of criticism for such an argument—the crucial ways out—are limited by this drawback.

The reconstruction represented in diagram 1 contains elements that it should not contain. Diagram 2 is, on the contrary, too concise. I will, therefore, present my own compromise reconstruction. What are the requirements that a successful reconstruction of Galileo's experiment should meet? First, the condition of *generalization* must be met: the reconstruction must not contain premises with particular details, as those distinguish straightforward arguments from experiments. If we want to defend the view that the absence of details does not affect the epistemic power, we must do without them, of course. The reconstruction must also be *adequate*: the straightforward argument must be an instance of the same scheme of reasoning as the thought experiment. Premises and the way the conclusion is derived from them are argument identifiers.

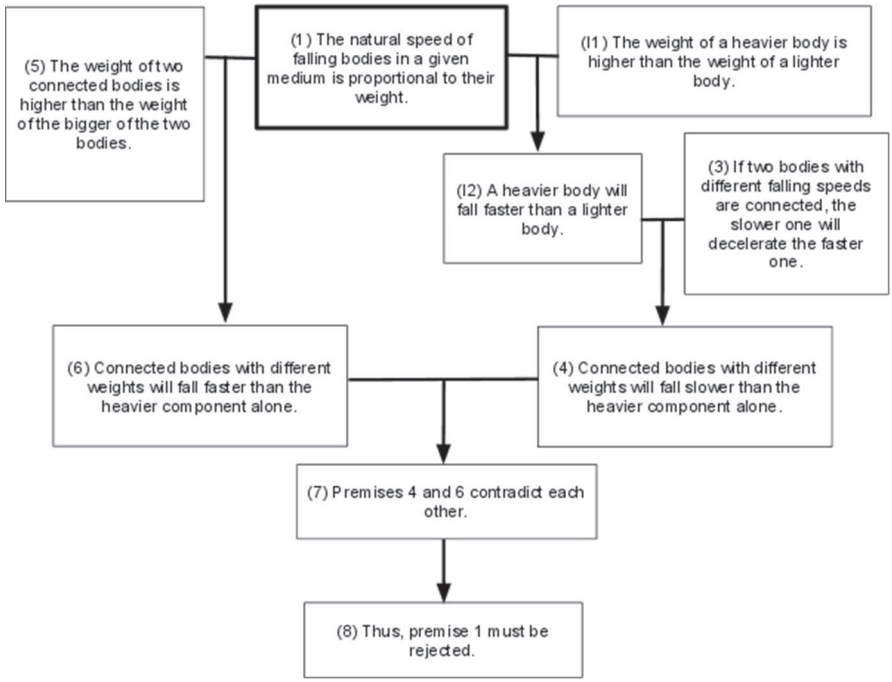

Figure 7-2: A compromise reconstruction

They tell us whether the straightforward argument is a reconstruction of the thought experiment or whether it is a different, independent argument. Third, there is a condition of *plausibility*: the reconstruction must work with premises whose plausibility, which is responsible for blocking the ways out, is the same as the plausibility of the relevant premises of the thought experiment. Gendler, of course, believes that the third condition (plausibility) cannot be met due to the first condition (generalization). The plausibility of the generally stated premises in her minimal reconstruction is really incomparable with the plausibility of the particular premises of the experiment. Could we perhaps find statements that are both general and plausible?

Diagram 3 offers such a reconstruction of Galileo's example. The condition of generalization is met: the argument does not contain any premises giving particular weights and speeds of bodies. It does not mention stones weighing so and so much, falling at such and such speed. It only mentions a heavier and a lighter body, a connection of such bodies, and their relative weights and speeds. The adequacy condition is met as well, the reductio ad absurdum is clear: the Aristotelian assumption is present at the beginning of both argument branches. Its role in both branches is the same—it is a rule that, when combined with the identification of object weight, determines its natural speed. The reconstruction contains two sub-arguments that lead to incompatible results from a common premise. In one branch, the Aristotelian principle of natural speed is applied to bodies before connection. In the other, it is applied to them after their connection. The satisfaction of the third condition will be demonstrated using the Toulmin model of argument in the following subsection.

4.1 Reconstruction: a functional analysis

The diagrams presented above may be clearer than a simple list of premises and conclusions—they even contain additional information about the relations between the individual premises—but they still do not exhaust everything that can be represented in a reconstruction. For a further analysis, I will use a stronger analytic tool—the Toulmin model—which can represent a characteristic that I consider crucial for the refutation of Gendler's critique. I repeat that the core of Gendler's critique is the assertion that the straightforward argument does not lead to the conclusion in the same way as the thought experiment. Gendler believes that the argument cannot only be refuted by denying the Aristotelian principle, but by other means as well. Also, the denial of the principle has no privileged status in the argument. The reason for that is that while we accept certain premises

in conducting the thought experiment, these are subject to independent evaluation in the argument. In particular, the ways out identified above are prima facie blocked in the thought experiment, but we need additional, nontrivial premises to block them in the argument.

Using the Toulmin model, I will attempt to show that the alleged difference between the experiment and its reconstruction does not exist, i.e. the ways out are equally blocked in both versions. The Toulmin model distinguishes premises on the basis of their argumentative function. The simplest model uses two types: data and warrants. Data are statements that indicate *"facts and present them as the foundation upon which our claim is based"* (Toulmin 1958, 90). It is the kind of premises that provide the needed construction material to support the conclusion, i.e. they provide the initial information about facts, events, individuals, and other things. The data of a given argument represent the initial points of support. Toulmin characterizes them with the question: *What have you got to go on?* Warrants are statements that say how the data is connected with the conclusion. They describe the steps needed for the data to support the conclusion. A warrant does not offer additional factual information or additional evidence. It shows how the factual information argumentatively relates to the conclusion. Toulmin characterizes warrants by means of the question: *How do you get there?* The warrant is sometimes described as an inferential license that relates the data to the conclusion. For clarity, here is an example of an argument with obvious functions for the premises:

> [a datum] Peter is a librarian—[the warrant] Librarians can read—[the conclusion] Peter can read.

Doubts have been expressed about some aspects of the Toulmin model, even at this elementary level. Some critics[8] point out that the criteria of identification of data and warrants are rather vague. In everyday argumentation, it is sometimes difficult to distinguish which statements are facts and which concern the way that the facts relate to the conclusion. It may not be always clear whether the critic of an argument requires the addition of data, or whether she is trying to identify the warrant that justifies the move from the data to the conclusion. How are the categories to be applied in arguments whose premises do not concern facts, such as hypothetical syllogisms? Toulmin was aware of some vagueness in the definition of data and warrants, and admitted that there were situations in which it was impossible to determine the function of a statement uniquely.

[8] See, for instance, Freeman (1991, 51) or van Eemeren, Grootendorst, and Kruiger (1984, 205).

He stressed, though, that his objective was not to provide precise terminology and demarcation, but to show that it was possible to categorize the premises of an argument very well in some contexts and use the categorization in the analysis of a dispute. I need not defend the universal character of the Toulmin model here. I need not presuppose that the function of a statement in an argument is always clear. It is sufficient for my purposes to accept the central idea of the Toulmin model, i.e. the idea that, at least in some contexts, the premises can have different functions relatively to the conclusion and can either provide facts or relate the facts to the conclusion. Using Toulmin's functional distinction, I propose to model the compromise reconstruction in Diagram 4.

The diagram represents both branches of the argument. Its conclusion, which states the incompatibility of 4 and 6, is not important for our purposes. The argument keeps the form of a reductio ad absurdum in which the initial accepted premise is the Aristotelian premise 1. That premise leads to incompatible conclusions after the application of two different principles. The argument contains an implicit, analytically true principle I1, which leads to the intermediate conclusion I2 in combination with 1. Premises 3 and 5 are presented as warrants that are prima facie plausible. After a more careful consideration it can, of course, turn out that the premises are false and the argument is unsound, but the same holds for the thought experiment. Premises 3 and 5 are as fallible as the particular premises of the thought experiment.

I have also included the metaphysical thesis D1 in the model. What is its argumentative function? One of the distinguishing properties of the warrant, according to Toulmin, is that it is usually supported by another statement that he calls the backing (Toulmin 1958, 91-92). Its purpose is to define the conditions for the application of the warrant. The backing in my argument is represented by the two general metaphysical theses D1 and D2, whose role in the argumentative reconstruction is to block the ways out. Both theses define the conditions for the application of warrants 3 and 5: thesis D1 says that the principles apply invariably and there is no exception for unified objects; thesis D2 says there is no exception with respect to the rate of unification of objects.

I reconstruct the metaphysical theses D1 and D2 as premises of sub-arguments whose conclusions are warrants 3 and 5. Functionally, they are the backing and their role is to support and determine the conditions of validity of the relevant warrants. Explicit statement of these supplementary premises/backings/theses may lead to the refutation of some objections. In particular, the explication of D1 leads to the refutation of objections C1 and C2, and similarly for D2, which blocks the remaining ways out.

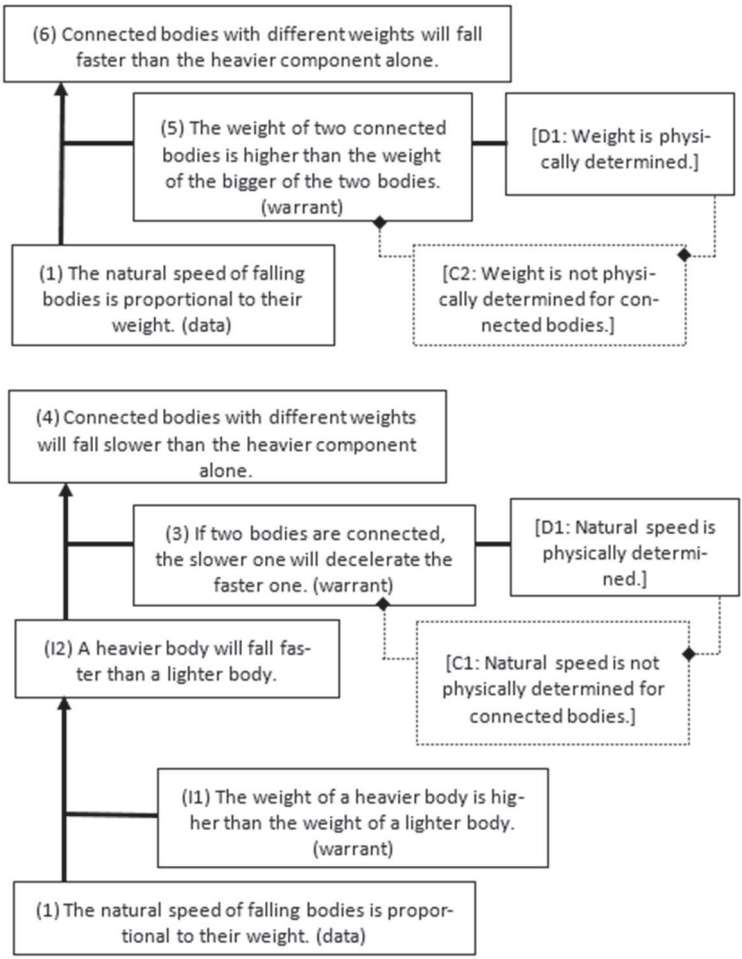

Figure 7-4: A partial Toulmin model of Galileo's example

4.2 Support vs. justification

I repeat the essence of Gendler's argument. Explicitly stated metaphysical theses are not part of the thought experiment in Galileo's example, simply because they are not needed. The straightforward reconstruction must, however, rely on some warrant backing due to possible objections, and, as a result, controversial metaphysical theses are added to the reconstruction.

My objection to Gendler's argument is that she incorrectly identifies the argumentative function with the epistemological function. The metaphysical theses surely have argumentative function. Being the backing, they support the warrants and offer reasons to accept them. That does not mean that the theses give the actual reasons why the warrants are accepted. The fact that X is a reason for accepting Y in the argumentative sense does not mean that Y has been accepted for reason X in the epistemic sense.

Let me illustrate the point. How did you learn about the Pythagorean Theorem? I assume you were told by your parents or teachers when you were young. Suppose you accepted the belief on the basis of the testimony of a reliable source. It was only later that you were introduced to the rules of mathematical proof that helped you demonstrate the premises that the Pythagorean Theorem depends on and the way it can be deduced from them. In the epistemic sense, the main reason for accepting the theorem is the authority of the testimony. The plausibility is further increased later by the deduction. The deduction is not the main epistemic reason for accepting the theorem; the theorem is accepted on the basis of another epistemically valid procedure. The same holds for beliefs about the physical world. We find out about the existence of regularities in the world by means of our senses. But we accept the existence of physical laws either on the basis of our own reasoning or on the basis of a testimonial. Sometimes we are able to deduce the laws from basic premises, sometimes not. But the point is that deduction is not the only or the primary means of obtaining and justifying beliefs.

By explicating D1 and D2, Gendler has shown a good way to deduce the key principles from general premises. However, it is only one of the ways that the principles can be epistemically justified. I believe that principles 3 and 5 are actually obtained by a different method: induction. We derive, generalize, accept and, if needed, explicate the principles on the basis of our experience with the operation of the world and the behaviour of various objects under various circumstances. We accept principles 3 and 5 not because we have deduced them from general premises, but because we derived them from particular situations in early childhood. We repeatedly observed an increase in weight after adding one weight to another. We repeatedly felt deceleration when we held the hand of a slower runner. That is why we obtained the beliefs that two things weigh more than one of them and that the resulting speed of connected objects will be slower than the speed of the faster one of them. When we evaluate Galileo's thought experiment and its argumentative reconstruction, we assume the truth of these beliefs and do not question their prima facie plausibility. If these warrants are attacked, we can, secunda facie, support them by some backing: we can refer to general and common experience with the behaviour

of bodies or offer a type of derivation from some more general premises. None of the ways of backing are necessarily related to thought experiments or straightforward arguments.

I claim that metaphysical theses D1 and D2 provide argumentative support for the warrants, but do not correspond to the process of their epistemic justification. The theses are not the reasons why the warrants are prima facie accepted, nor are they needed to refute the ways out. They only become relevant if we want to refute the ways out by a special method, namely by a deductive relation of the attacked warrants to elementary metaphysical premises.

4.3 Details

The analysis has shown which premises are presented as starting points in the argument and what their functions are. Finally, we must answer the question of the role of the particular details. According to Gendler, they are responsible for blocking the ways out, i.e. they have an argumentative function. The given statements provide the backing due to which some objections are refuted. There is a relation of argumentative support between the statements and the warrants and a reconstruction without them would be incomplete. The support of the warrants would not be fully expressed without the particular premises, as the argument would be open to some objections, unlike the thought experiment.

Is the role of the particular statements in Galileo's experiment really as Gendler claims? Is there a relation of support between the particular statements and the warrants that needs to be included in the reconstruction? I do not think so. The particular statements serve a different function; their task is not to persuade us about the validity of a general principle, but to help us understand it. It is a difference in the interpretation of the argumentative role of a particular example, which Perelman and Olbrechts-Tyteca call a difference between an *example* and an *illustration* (Pereman and Olbrechts-Tyteca 1958, 350 and further). An example is a way of using a case in which the particular statement is meant as a reason for generalization. A particular statement that is an example thus precedes generalization; a general statement is obtained only from a particular example. An illustration is a different way of using a particular case—a general statement is not derived from it, but it is presupposed. Perelman and Olbrechts-Tyteca describe illustration as a particular case whose role is "*to strengthen adherence to a known and accepted rule"*. An illustration, according to them, "*clarifies a general statement, shows the import of this*

statement by calling attention to its various possible applications, and increases its presence to consciousness".

The difference between an example and an illustration is analogical to the difference between an argument and an explanation. The latter difference is also characterized by the dialectical status of a thesis. While the thesis of an argument (conclusion) is a point of dispute and must be supported, the thesis of an explanation (explanandum) is accepted by the parties concerned and is not in itself controversial. It is not always easy to determine whether a given statement is meant as an argument or an explanation. Neither is it always clear whether a particular statement is presented to derive a generalization or to illuminate a generalization that has already been introduced. However, this distinction is important for the evaluation. An unsuccessful example means that the proponent has failed to support her thesis and the audience has no reason to accept it; there is, thus, a dialectical consequence. An unsuccessful illustration means that the proponent has failed to strengthen a thesis that the audience has already accepted; the consequence is rhetorical.

Walton (2008, 314) describes the scheme of an argument based on example as follows: An individual has the property P and the property Q in a particular case. Thus, it is generally true that if x has P, x also has Q. He identifies five questions, the answers to which are needed for the quality of the argument. One of the questions is: Does the cited example support the generalization, or is it an instance of it? The answer to this question is essential to distinguish whether a particular case is meant as an example or whether it is an illustration. Let us return to Galileo's thought experiment. The particular statements in his example are not meant to persuade us that natural speed is mediative and that the weight of the whole is higher than the weight of its parts. These principles are deeply rooted in our understanding of the physical operation of the world, and we need not be persuaded by them. The particular statements help us understand that these general and accepted principles of the movement and weight of bodies also apply to free fall, and, therefore, are not compatible with another relevant Aristotelian assumption. Thus, I claim that the particular details are used as illustrations in the thought experiment.

The stones of particular weight falling at particular speed in Galileo's thought experiment are presented as situations instantiating general principles regarding their combined weight and speed that have already been accepted. The goal of the particular example is to point out that the inductively obtained principles are valid even in cases of objects falling freely. The particular premises do not support the plausibility of warrants 3 and 5 argumentatively, and, thus, have no epistemic power. The generalization

of Galileo's thought experiment, therefore, has no effect on its epistemic value.

Conclusion

Gendler describes the crucial place in the polemic about eliminativism as follows:

> Contemplation of the case Galileo describes *brings him* [the proponent of Aristotelian physics, MP's note] *to see* that these principles are not defeated in *this* case. And it is this recognition that serves as the basis for the case's power. No austere argumentative reconstruction will be able to do this, because part of the thought experiment's function is to bring the Aristotelian to accept certain *premises*. (Gendler 1998, 408)

The experiment with particular details, in her opinion, makes the recipient accept certain principles and thus excludes some possibilities for criticism. The straightforward argument does not enable this and all possibilities are open if they are not blocked by further, controversial premises. I have shown that this opinion is based on an inadequate reconstruction of Galileo's thought experiment. The alleged difference between the experiment and the argument is illusory. The reconstruction offered by Gendler contains principles that are formulated very generally and that are most plausibly justified by deductive relations to even more general and problematic premises. This illusion disappears in my reconstruction. I have based the straightforward argumentative reconstruction of the experiment on general but highly plausible principles where the ways out are blocked with the same strength as in their imaginary particular counterparts. The key premises are as plausible as the key premises of the thought experiment, i.e. the argumentative reconstruction makes the recipient accept the premises as much as the thought experiment. Their justification is based on inductive generalizations of past experience.

Enthusiasts believe that the particular details in the interpretation of Galileo's experiment have an argumentative value. The particular case is seen as an example that supports the validity of the general principles. This interpretation is, however, not adequate in the context of Galileo's example. It is more plausible to interpret the particular details as illustrations. That way their argumentative relation to the general premises, which is a necessary condition for an epistemic relation, disappears. The particular details then cannot have an epistemic function as they lack an argumentative function.

Gendler claims that the principles in the thought experiment are validated by the particular cases. I have presented evidence that the principles have already been accepted on inductive grounds, their validity is assumed, and their point is merely illustrated by the particular examples. In my argumentative analysis, I have drawn attention to three mistakes that Gendler makes when criticizing eliminativism: (#1) her demands in the generalization of the thought experiment are too high; (#2) she incorrectly identifies argumentative support of the premises with their epistemic support; (#3) she incorrectly identifies the argumentative function of the particular cases.

Finally, I would like to state the goal of this paper more precisely. It is to defend eliminativism against the criticism relying on Galileo's great thought experiment. I have tried to show that this particular experiment can be reconstructed without a loss of epistemological value, which is directly opposed to enthusiasm. I do not claim, though, that *all* thought experiments can be reconstructed without such a loss.

Bibliography

Bishop, M. 1999. “Why Thought Experiments Are Not Arguments”. *Philosophy of Science*, Vol. 66: 534-541.

Borsboom, Denny.et al. 2002. “Functional Thought Experiments”. *Synthese*, Vol. 130: 379-387.

Brown, James Robert. 1991a. Laboratory of the Mind: Thought Experiments in the Natural Sciences. London: Routledge.

Brown, James Robert. 1991b. “Thought Experiments: A Platonic Account”. In *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and G. Massey, 119-128. Lanham: Rowman & Littlefield.

Camp, Elisabeth. 2009. “Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments”. *Midwest Studies in Philosophy*, Vol. 33, No. 1: 107-130.

Freeman, James B. 1991. Dialectics and the Macrostructure of Arguments. Berlin: Foris.

Galilei, Galileo. 1974. *Two New Sciences*, translated by S. Drake. Madison: University of Wisconsin.

Gendler, Tamar Szabó. 1998. “Galileo and the Indispensability of Scientific Thought Experiment”. *The British Journal for the Philosophy of Science*, Vol. 49, 397-424.

Häggqvist, Sören. 2009. “A Model for Thought Experiments”. *Canadian Journal of Philosophy*, Vol. 39: 56-76.

Mach, Ernst. 1905. “Über Gedankenexperimente”. In *Erkenntnis und Irrtum*, 181-197. Leipzig: Verlag von Johann Ambrosius Barth,

Mach, Ernst. 1960. The Science of Mechanics: A Critical and Historical Account of its Development. La Salle: Open Court.

Nersessian, Nancy. 1993. “In the Theoretician’s Laboratory: Thought Experimenting as Mental Modeling”. *Proceedings of the Philosophy of Science*, Vol. 2, 291-301.

Norton, John D. 1996. “Are Thought Experiments Just What You Thought?”. *Canadian Journal of Philosophy*, Vol. 26: 333-366.

Norton, John D. 2004a. “Why Thought Experiments Do Not Transcend Empiricism”. In *Contemporary Debates in the Philosophy of Science*, edited by C. Hitchcock, 44-66. Oxford: Blackwell.

Norton, John D. 2004b. “On Thought Experiments: Is There More to the Argument?”. *Philosophy of Science*, Vol. 71: 1139–1151.

Perelman, Chaïm and Olbrechts-Tyteca, Lucie. 1958. *La Nouvelle Rhétorique: Traité de l’Argumentation*. Paris: Presses Universitaires de France.

Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.
Toulmin, Stephen. 1958. *The Uses of Argument*. Cambridge: Cambridge University Press.
van Eemeren, Frans. Hendrik, Grootendorst, Rob, and Kruiger, Tjark. 1984. *The Study of Argumentation.* New York: Irvington.
Walton, Douglas. 2008. *Argumentation Schemes*. Cambridge: Cambridge University Press.

CHAPTER EIGHT

"NEWTON'S GENESIS": A THOUGHT EXPERIMENT ON THE CREATION OF "BODIES" AND ITS EPISTEMOLOGICAL CONSEQUENCES[1]

MATJAŽ VESEL

Introduction

The natural philosophy of the Scientific Revolution and the scientific discourse of the following centuries are heavily populated by thought experiments (see McAllister 1996; 2005; 2013[2]). Even Isaac Newton, the most famous representative of experimental philosophy, is credited with several of them. The two best-known, the bucket experiment and the rotating spheres experiment, are from scholium to definitions of *Principia Mathematica Philosophiae Naturalis* (1687/1999, 68),[3] but there are many others.[4] Consider, for instance, the anecdote related by Pierre Coste, the translator of John Locke's *Essay on Human Understanding* into French, in the Amsterdam edition of the translation. In one of the footnotes, Coste reports on Newton's conversation with the author of the *Essay*, John Locke,

[1] This article is a result of the research program P6-0014, "Conditions and Problems of Contemporary Philosophy", and the research project J6-9392, "The problem of objectivity and fiction in contemporary philosophy", which are funded by the Slovenian Research Agency.

[2] It is questionable, though, whether all experiments discussed by McAllister are genuine thought experiments. For an excellent discussion on what qualifies as a genuine thought experiment, see Roux 2011. I find her distinction between an idealized experiment and a genuine thought experiment especially important.

[3] On these thought experiments, see, e.g. DiSalle 2016 and Laymon 1978. See, however, the above note.

[4] Some additional ones from *Principia* are mentioned by Harper 2016 and Pourciau 2016. The same reservation as expressed in previous two notes applies here as well.

and Thomas Herbert, at that time the president of the Royal Society, regarding the question of how God created matter. According to Coste's report, Newton wanted them to suppose that God, first, made some part of necessary, infinite, eternal, and penetrable "pure space" impenetrable, so "that portion of space would possess impenetrability, which is one of the essential qualities of matter" (Bennet and Remnant 1978, 5)[5] and, second, that he communicated impenetrability to another similar portion of space, which should provide them with "some idea of the mobility of matter, another quality which is also utterly essential to it" (Bennet and Remnant 1978, 5).

Coste's account suggests not only that Newton considered impenetrability and mobility to be two qualities of matter, but also that he believed them to be its *essential* qualities: matter is what it is because of its impenetrability and mobility. In other words, according to Coste, Newton believed that the essence of matter—or the essence of body, since matter and body are synonyms for him—can be defined as an impenetrable and mobile part of space. In the light of the notorious lack of any definition of the nature of body in Newton's *Principia mathematica philosophiae naturalis*, this is rather strange.[6] In the *Principia*, Newton does not attempt to define the essence or nature of the body, but settles—for not very manifest reasons, at least not at first sight—for an understanding of bodies at the phenomenal level. In "General Scholium" he affirms the following:

> We see only the shapes and colors of bodies, we hear only their sounds, we touch only their external surfaces, we smell only their odors, and we taste only their flavors. But there is no direct sense and there are no indirect reflected actions by which we know innermost substances. (Newton 1687/1999, 942)

Coste's account—assuming that his relation of the conversation is more or less accurate—raises several questions. Why did Newton and his interlocutors bother with the question of the genesis of matter at all? And why did Newton propose a thought experiment to resolve the question? Would it not be more adequate, natural, so to speak, and in line with Newton's experimental philosophy, to deduce, possibly via some experiences,

[5] English translation of Locke (1735, 521) by Bennett and Remnant 1978.

[6] See Jalobeanu 2007 and 2013. For a broader context of the discussion on the nature of body/matter, see e.g. Garber, Henry, Joy and Gabbey 1998; Jalobeanu 2011; and Damerow 2016. I will have to leave aside many important aspects of Newton's philosophy of matter/body. For discussions on some of them, see e.g. McGuire 1982; Janiak 2008; Schliesser 2010; Biener and Smeenk 2012; Brading 2012; and Palmerino 2013.

observations, or experiments, the manner of creation of matter from the existing, actual matter in the universe? What are we supposed to learn from this thought experiment? And how does the definition of the nature of matter/body as a mobile and impenetrable part of space square with *Principia*'s lack of any such definition? In order to answer these questions, we need to contextualize them, and we need to do so on several levels.

Newton's suggestion, as narrated by Cotes, is actually a very condensed and not completely accurate summary of Newton's account of God's possible creation of "bodies"[7] (i.e. matter) developed in one of his unfinished manuscripts, for a long time unknown to scholars, which was edited and published for the first time in 1962 by A. Rupert Hall and Marie Boas Hall, and now known under the rather misleading title of *De gravitatione*.[8] *De gravitatione* was conceived as a text on hydrostatics,[9] but Newton soon digresses into a critical philosophical debate with Cartesian natural philosophy and metaphysics,[10] criticizing Descartes's notion of motion, his conception of body as being identical to extension/space and what seems to follow from this identification, and then returns to hydrostatic issues. In this critical process, Newton develops his own notions of motion, space, body, and related notions, including that of God's role in the operations of the physical world, and claims that his theory of body, being distinct and independent of space/extension, *confirms* and *explains* the principal metaphysical truths.

Newton's Genesis as a Thought Experiment

While Newton makes some brief remarks on body during his presentation of the alternative notion of "space alone"—necessary, infinite in all directions,

[7] I am putting "body" and "bodies" in brackets in order to underline that in *De gravitatione*, as we shall see shortly, Newton does not discuss bodies but beings in every way similar to bodies. Jalobeanu (2007), calls them "pseudo-bodies".

[8] See Newton 1962, 89–121 (Latin text) and 121-156 (English translation). For the amended translation by Christian Johnson with the help of Andrew Janiak, see Newton 2014, 26–58. All my quotes from *De gravitatione* are from this translation.

[9] The first sentence in the original reads: "De gravitatione et aequipondio fluidorum et solidorum in fluidis scientiam duplici methodo tradere convenit" (Newton 1962, 90).

[10] Newton studied Descartes' works, especially *Principles of Philosophy*, very thoroughly. The best testimony to it is *De gravitatione* itself, but traces of his detailed study of Descartes are very clear also in his "Qaestiones quaedam philosophicae" from the *Trinity College Notebook* (Newton 1661-65). See McGuire and Tamny (Eds.) 1983 and the editors' comprehensive introduction. Newton was also familiar with Descartes' famous fable du monde, but only in a summary given in *Discourse on Method*.

Euclidean, eternal, immobile, homogeneous, and independent of bodies (Newton 2014, 35-41)—he dedicates a separate portion of the text specifically to the "nature of body" or the problem of the distinction between space/extension and body, which can be divided into two parts. In the second part of this section he attacks Descartes's thought experiment of receding bodies (*Principles of Philosophy*, II, 4 and 11) and shows—at least in his own opinion—that Descartes's reasoning is faulty (Newton 2014, 48-50). It is the first part, however, where he elaborates on God's creation of "bodies," that will concern us. (Newton 2014, 41-48)

Somewhere in the middle of *De gravitatione*, Newton elaborates on what would follow if God used his power to create "a certain kind of being similar in every way to bodies, and whose creation we cannot deny to be within the power of God" (Newton 2014, 42), in order to elucidate his understanding of the nature of body in contrast to that of Descartes', and describes the possible creation of matter by God in a similar way as in the abovementioned conversation, adding some further elaborations. Newton develops his line of reasoning with reference to God's free power to move bodies at will and to the fact that he can by "the sole action of thinking and willing" (Newton 2014, 42) prevent "a body from penetrating any space defined by certain limits" (Newton 2014, 42). The first inference follows, according to Newton, from our, i.e. human, ability to move our biological bodies with thought/will alone: if we can do it, God, "whose faculty of thought is infinitely greater and more swift [than ours]" (Newton 2014, 42), can do it as well. The second inference simply follows from the infinite power of his will: while we can move our bodies by will, we obviously cannot create bodies by will and cause them to be impenetrable. But this is, for the abovementioned reason, obviously within God's power.

Then Newton asks what would happen if God were to exercise this power and "cause some space projecting above the earth, like a mountain or any other body, to be impervious to bodies and thus stop or reflect light and all impinging things?" (Newton 2014, 42) Our senses—sole judges in this matter, according to Newton—would persuade us to believe that this space is a real body. This space in the shape of a mountain would be regarded as tangible, because of its impenetrability, "and visible, opaque, and colored on account of the reflection of light, and it will resonate when struck because the adjacent air will be moved by the blow" (Newton 2014, 42). In the same manner, we may suppose that there are many empty spaces in the world, "one of which, defined by certain limits, happens by divine power to be impervious to bodies" (Newton 2014, 42). By the same hypothesis, it is also clear that this space would resist the motions of bodies and perhaps reflect them and assume all the properties of corporeal particles, except that it

would be regarded as motionless. The next step follows: mobility. If we suppose further that this impenetrable shaped space could be "transferred here and there according to certain laws" (Newton 2014, 42)[11], without any changes in its impenetrability and shape, it would have all the properties of body: shape, tangibility, mobility, ability to reflect and be reflected, it could constitute the structure of things, and it could operate on our minds and be acted upon, "because it would be nothing other than the effect of the divine mind produced in a definite quantity of space. For it is certain that God can stimulate our perception by means of his own will, and thence apply such power to the effects of his will" (Newton 2014, 42). And finally the last step: universalization. Let us further suppose other spaces of this kind, impervious to bodies and each other: "they would all sustain the vicissitudes of corpuscles and exhibit the same phenomena" (Newton 2014, 43). The conclusion is more than evident: "And so if all of this world were constituted out of these beings, it would hardly seem to be inhabited differently. And hence these beings will either be bodies, or very similar to bodies" (Newton 2014, 43).

The conclusion Newton draws from his thought experiment on the creation of matter is not that these beings *are* bodies. They could be, but they could also be just "very similar to bodies". Although it seems that Newton's bet is on them being bodies, this is not what he explicitly claims.[12] *If* they are bodies, he says, *then* the bodies can be defined as "*determined quantities of extension which omnipresent God endows with certain conditions*" (Newton 2014, 43).[13] mobility, impenetrability, and ability to excite perceptions of the senses and in the imagination in created minds, and conversely be moved by them.[14]

In the commentary that follows, Newton elaborates on six philosophical points—essential by-products—of his conception of body as distinct from space/extension. I shall mention only the last two. The fifth point is that a description of the corporeal body has been deduced "from our faculty of moving our bodies, so that all the difficulties of the conception may at length be reduced to that" (Newton 2014, 45). And the last one concerns the

[11] Note that Newton does not specify the laws in question.

[12] This seems to be the position of Martin Tamny; he writes: "Although this theory is presented at first as only a possibility, it quickly becomes clear that Newton believes in its truth" (Tammy 1979, 50).

[13] The manuscript has larger characters that are replaced with italics in the edition of Hall and Boas Hall and in the English translation in Newton 2014.

[14] For different, sometimes conflicting interpretations on how these three conditions are supposed to be properly understood, and some related issues, see, for instance Dempsey 2006, 2009 and 2014; Slowik 2009; Gorham 2011a and 2011b; Kochiras 2011 and 2013; Henry 2014; and Stein 2016.

usefulness of the idea of the described body. This can be seen in the fact that "it clearly involves the principal truths of metaphysics and thoroughly confirms and explains them" ((Newton 2014, 45).

It is, I believe, obvious from the above summary that we are here dealing with an exemplary thought experiment.[15] Newton's genesis is counterfactual, it involves a concrete scenario, and it has a well-delimited cognitive intention.

As we have seen, Newton does not discuss the nature of the real, actual bodies in the world, but describes some kind of being that is in every way similar to bodies. Newton's genesis is not conducted in reality, but is achieved in thought alone, through a series of suppositions. This brings us to the second characteristic of a thought experiment. It presents a vivid specific case, "a scenario". In our case it has several steps. First there is a concrete example: let us suppose that God causes some space above the Earth in the shape of a mountain, for instance, to be impervious to bodies. Second: now that we see, by this concrete example, what we are up against, we may suppose that there are many empty spaces in the world and one of them is made impenetrable. Thus it will assume all the properties of corporal particles with the exception of mobility. Next, we should suppose that this one impenetrable space is moved, thus adding mobility to its properties. Finally, we may suppose that God made more spaces impenetrable and mobile. The result of our thought experiment is the material universe of bodies as we perceive it. Newton's genesis also satisfies the third criterion: it has well-defined cognitive intention. Newton's genesis is designed "for framing a specific, preconceived thought goal" (Roux 2011, 24), to use Roux's words, it clearly delimits "a before and an after in what we know and what we think" (Roux 2011, 25). Newton's genesis is supposed to let us know that there is, contrary to what Descartes believes, a clear-cut distinction between the nature of body and the nature of extension/space:

> So much for the nature of bodies, which in explicating I judge that I have sufficiently proved that such a creation as I have expounded is most clearly the work of God, and that if this world were not constituted from that creation, at least another very like it could be constituted. And since there is no difference between the materials as regards their properties and nature, but only in the method by which God created one and the other, the distinction between body and extension is certainly brought to light from this. (Newton 2014, 47-48)

[15] Slowik, in "Newton's Metaphysics of Space" (2009), calls Newton's argument the "determined quantities of extension hypothesis." Stein, in "Newton's Metaphysics" (2016), the "creation story" or the "creation hypothesis", but Dempsey, in "Newtonian Idealism" (2014), sometimes refers to it as a thought experiment.

Thus we have the answer to the third question that I posed at the beginning of this article: What are we supposed to learn from this particular thought experiment? The distinction between body and extension. Now we can turn to the remaining questions: Why does Newton feel obliged to discuss God's creation of matter? Why does he, as the paradigmatic experimental philosopher, not examine the actual state of the matter in order to deduce how it was created? These two questions seem to be even more pertinent since Newton, in the draft related to one of the queries of *The Opticks*, affirms something that seems to be directly contrary to them. In this draft he claims that "[t]he business of Experimental Philosophy is to find out by experience & observation not how things were created but what is the present frame of Nature" (Newton 1675, fol. 242^{v}).[16] And, finally: What are the consequences of Newton's thought experiment regarding the question of whether the body can be defined?

Theological Voluntarism

Newton's declared reason for overthrowing Descartes's philosophy as regards the identification of body with extension is "to lay truer foundations of the mechanical sciences" (Newton 2014, 34). But he has other goals in mind as well. Within the section on "the nature of body", he reveals deeper and broader—both theological and metaphysical—reasons for his attack on Descartes. In the commentary on his theory of body, he explains that his description of the corporeal body was deduced from our ability to move our bodies in order "that God may appear (to our innermost consciousness) to have created the world solely by the act of will, just as we move our bodies by an act of will alone" (Newton 2014, 44). Even more, it was deduced from our ability to move our bodies in order "to show that the analogy between the divine faculties and our own may be shown to be greater than has formerly been perceived by philosophers" (Newton 2014, 44). Newton justifies his view with the analogy between the divine faculties and our own with the theological authority of the Bible: we were created in God's image, which would be even more visible "if only he simulated in the faculties granted to us the power of creation in the same degree as his other attributes."[17]

---

[16] See also fol. 243^{r}.

[17] *Ibid*. God's creating power is actually rather poorly delineated in our ability to move our bodies. By moving them we do not create anything, nor can we create anything, but rather only, "simulate the power of creation" (Newton 2014, 44). And even in that case there are severe limitations. We cannot make any part of space impenetrable to bodies; we only move bodies. We cannot move bodies of our

But even more significant is Newton's explanation of the benefits of his conception of "body" as completely dependent upon God's will. The usefulness of his notion of "body" can be seen in the fact that "it clearly involves the principal truths of metaphysics and thoroughly confirms and explains them" (Newton 2014, 45). These principal truths of metaphysics are: the existence of God; God's creation of bodies in empty space *ex nihilo*; and the distinction between these beings and created minds on the one hand, and their ability to unite on the other. Descartes' identification of extension/space and body is a sure path to atheism for two reasons. In Descartes's philosophy, extension, which for him is identical with body, "is not created but has existed eternally" (Newton 2014, 46), and "because we have an idea of it without any relation to God, and so in some circumstances it would be possible for us to conceive of extension [= body] while supposing God not to exist" (Newton 2014, 46). We are now getting to the crux of the matter: Newton's principal goal in developing his theory of body as distinct from space/extension is to avoid the pitfalls of atheism.[18]

In order to create natural philosophy without atheistic potential, Newton has to rely on his own version of what is nowadays usually labeled theological voluntarism.[19] It is constituted, to put it simply, by two closely

---

choosing, but only our own bodies to which we are united. We are not united to our bodies by our own will, but by divine constitution. We also cannot move bodies completely at will, but only in accord with those laws that God has imposed on us. Newton agrees – he concedes that this power could be called, "the finite and lowest level of the power which makes God the creator" (Newton 2014, 45) –, but still insists that it reflects God's creative power, i.e. that it does not detract from the divine power more than it detracts from his intellect that intellect belongs to us in a finite degree.

[18] See also Craig 1727:

> I shall not tell you what great improvements he [i.e. Newton] made in Geometry & Algebra, by which he was enabled to finish the two forementionʼd Books. But it is proper to acquaint you that his great application in his inquirys into Nature did not make him unmindfull of the great Author of Nature; they were little acquainted with him, who imagine that he was so intent upon his studys of Geometry & Philosophy as to neglect that of Religion & other things subservient to it. I am very {illeg} \And this I know/ that he was much more fellicitous in his inquirys into Religion than into Natural Philosophy; & that the reason of his showing the errors of Cartes's Philosophy, was because he thought it was made on purpose to be the foundation of infidelity.

There is a lot of studies on the reception of Cartesian philosophy in Britain and on how it was perceived as an atheistic threat. See, for instance, Henry 2013.

[19] For accounts on Newton's theological voluntarism, see, for instance Oakley 1961; Heimann 1978; Davis 1911; and Dobbs 1991. Newton's theological voluntarism has

interwoven aspects: the first one is God's infinite, limitless power, and the second one is the analogy between God's power/will and human power/will.[20]

Newton's position regarding God and his creation of "bodies" from *De gravitatione* is very similar to that in one of his manuscripts from the early 1670s, entitled "Of natures obvious laws & processes in vegetation", and to "An Hypothesis explaining the Properties of Light", which Newton sent to the Royal Society in 1675.

In the first text, Newton explains his thoughts on God by referring to the principle that he can do whatever he wants provided it is not logically contradictory.[21] One of the examples he provides are the human faculty of knowledge and the power to activate matter (which could mean, in my opinion, the power humans have to move their own bodies, but the text is too vague to be certain) without limits. Since Newton can conceive his power of knowledge and power to activate matter as limitless without a contradiction, such powers either exist or may be made to exist. After these preliminaries, he finally addresses the issue of God, and even then not directly. He simply claims that the world "might have been otherwise then [= than] it is", because there is a possibility of worlds being designed differently. The world "as it is" is therefore not a result of a necessity, but of a free will, voluntary and free determination, and this free determination implies (the existence of) God. This is extremely important. Newton's main preoccupation seems to be—exactly the same as in *De gravitatione*—to

---

been challenged by Peter Harrison (2004). A successful case against his challenge was made by John Henry (2009). Harrison accepted Henry's criticism, except for some specifics. See his "Voluntarism and the Origins of Modern Science: A Reply to John Henry" (2009). See also Ducheyne 2012. For a more general view on Newton's theology and religion, see, for example, Iliffe 2017.

[20] On this, see Guerlac 1983 and Iliffe 1995.

[21] See Newton 2006, fol. 4[v].

> Of God. what ever I can conceive without a contradiction, either is or may bee made by something that is: I can conceive all my owne powers (knowledge, activating matter etc) without assigning them any limits Therefore such powers either are or may bee made to bee. /…/ Arg 2. The world might have been otherwise then it is (because there may be worlds otherwise framed then this) Twas therefore noe necessary but a voluntary & free determination that it should bee thus. And such a voluntary [cause must bee a God] determination implys a God. If it bee said the world could bee noe otherwise then tis because tis determined by an eternall series of causes, thats to pervert not to answer the 1st proposition: For I meane not that the Earth might have been otherwise notwithstanding the precedent series of causes, but that the whole series of causes might from eterity have beene otherwise here, because they may bee otherwise in other places[.]

See also Dobbs 1991: 266.

affirm the very existence of God, which depends, according to his reasoning, on the possibility of the creation of many different worlds.

The second text, "An Hypothesis explaining the Properties of Light" (See Birch 1757), is interesting because of Newton's stress on God bestowing self-motion on animals, which is beyond our understanding. We do not understand how animals, including human beings, move. By analogy, we do not know what the principle of the motion of light is: "God, who gave animals self-motion beyond our understanding, is, without doubt, able to implant other principles of motion in bodies, which we may understand as little" (Birch 1757).

If we put these two texts together, the conceptual cluster of *De gravitatione* appears. In this manuscript, the creation of matter is made dependant on God's free will, which means it could have been created differently. Newton builds his case in two steps. The first step concerns the creation of "bodies", and follows from God's power of creation and our—limited—understanding of his power. The result of the first step is negative: we cannot be sure whether we are dealing with beings that are metaphysically and essentially bodies or beings just completely similar to bodies. The second step involves the motion of these beings, which is inferred from our consciousness being able to move our biological bodies by will. Since we can do it, God can do this as well; even more, he can make these beings impenetrable.

In comparison with the nature of space, our explanation of the nature of body, claims Newton, "must be more uncertain, for it does not exist necessarily but by divine will, because it is hardly given to us to know the limits of the divine power" (Newton 2014, 41). The crucial problem is that we do not know the limits of the divine power and we consequently cannot know "whether matter could be created in one way only, or whether there are several ways by which different beings similar to bodies could be produced" (Newton 2014, 41). Newton states that it is highly unlikely "that God could create beings similar to bodies which display all their actions and exhibit all their phenomena, and yet would not be bodies in essential and metaphysical constitution" (Newton 2014, 41), but, he affirms it, since:

> as I have no clear and distinct perception of this matter I should not dare to affirm the contrary, and hence I am reluctant to say *positive* what the nature of bodies is, but I would rather describe a certain kind of being similar in every way to bodies, and whose creation we cannot deny to be within the

> power of God, so that we can hardly say that it is not body. (Newton 2014, 41-42)[22]

In the next step, Newton builds his argument on the analogy between human and God's will/thought and power. First, since all human beings can move their bodies at will, God, whose faculty of thought is infinitely greater, can do it as well:

> Since each man is conscious that he can move his body at will, and believes further that other men enjoy the same power of similarly moving their bodies by thought alone, the free power of moving bodies at will can by no means be denied to God, whose faculty of thought is infinitely greater and more swift. (Newton 2014, 42)

Second, since God's faculty of thought is infinitely greater and more swift than human, he should and is able not only to move "bodies" at will, but he can also "by the sole action of thinking and willing /…/ prevent a body from penetrating any space defined by certain limits" (Newton 2014, 42).

We are presented with uncertainty regarding the explanation of the nature of the body. This uncertainty is a consequence of our inability to grasp the limits of the divine power on the one hand, and of the nature of the bodies created by God's—we may infer—limitless, infinite power, on the other, which makes their existence non-necessary. While space exists necessarily—"space is an emanative effect of the first existing being" (Newton 2014, 40)—bodies do not, they exist solely because God decided that they should exist; they are effects of God's will, of his volition. But we do not know in how many ways matter, i.e. bodies, could have been created by God. Perhaps there is one way (i.e. the creation of actual bodies in the universe), or perhaps God could have created beings that are completely like bodies in many other ways. What puzzles Newton is, obviously, the second possibility: God could have created beings—not bodies!—that are similar to bodies, displaying *all their actions* and exhibiting *all their phenomena*, but still not being bodies in any essential and metaphysical sense. Although it is not very likely, it is still possible. According to Newton, God could

[22] I am leaving *positive* in Latin on purpose. Gabbey (2011, 429), proposes a translation of "*et proinde nolo positive dicere quaenam sit corporea natura*", which differs from that in Newton, *Philosophical Writings*: "and hence I am reluctant to say positively what the nature of bodies is". In Gabbey's view, it should be paraphrased as follows: "Accordingly I am unwilling to affirm anything about corporeal that is not evident to the senses and truly representative of a real essence." See his explanation on the same page, n. 15; see also the very end of this article.

have created—*mutatis mutandum*—a being that walks like a duck, swims like a duck, and quacks like a duck, but is not a duck—at least not in its *essential* and *metaphysical* constitution. God could have deceived us, in a sense. Since Newton does not have "a clear and distinctive idea" on the issue, he dares not declare that that is not the case, i.e. he cannot claim that God *did not* create beings which are in every aspect similar to bodies, except that they are not bodies metaphysically and essentially; therefore, he is unwilling, or one could say, he cannot provide a theory on the nature of body. The only thing that he can do is to *describe* a certain kind of being similar in every way to bodies, which can be created by God and which can only barely pass as "not a body".

Thought Experiment, Theological Voluntarism, and its Epistemological Consequences

Let us see whether my remaining questions can now be answered: Why have a discussion on how God could have created matter at all? How come Newton does not engage in an experimental and observational examination of "the present frame of Nature"? And what about the very possibility of defining body?

Newton's thought experiment on God's creation of matter/bodies is, partly, due to the need to establish a natural philosophy of body and space without atheistic potential.[23] Newton's "body" is not identical with space, and therefore does not exist eternally, but was created by God's free will. Hence, we cannot have an idea of "body" without any relation to God and we cannot conceive of "body" while supposing that God does not exist. But I would go even further and argue that Newton wants to beat Descartes with Descartes' own weapon; i.e. he wants to show that Descartes' own considerations regarding God's possibilities in the creation of the world, which Newton shares with him, lead to different specifics of natural philosophies, in Descartes' case to false and even potentially atheistic ones.

Descartes considers God's creation of the world on several occasions in *Principles of Philosophy* (2009),[24] but the crucial texts come at the end of the fourth part, where he addresses the question of how we may arrive at knowledge of the shapes, sizes, and motions of particles that cannot be perceived by the senses, and related issues (Descartes 2009, Part IV, art. 203–206, 288). The details of Descartes' explanation of his own path do not

[23] See also Henry 2009, 81-82.

[24] See also Curley 1993; Osler 1994; and Ariew 2005 (especially pages 137-138) and 2010.

need to concern as here; what is important is his conclusion. Although this method, he affirms, "may enable us to understand how all the things in nature could have arisen, it should not therefore be inferred that they were in fact made in this way" (Descartes 2009, Part IV, art. 204, 289). And he continues:

> Just as the same craftsman could make two clocks which tell the time equally well and look completely alike from the outside but have completely different assemblies of wheels inside, so the supreme craftsman of the real world could have produced all that we see in several different ways. I am very happy to admit this; and I shall think I have achieved enough provided only that what I have written is such as to correspond accurately with all the phenomena of nature. (Descartes 2009, Part IV, art. 204, 289)

Or, in the words of the title of article 204: "With regard to the things which cannot be perceived by the senses, it is enough to explain their possible nature, even though their actual nature may be different" (Descartes 2009, Part IV, art. 204, 289). Descartes is saying that God could have produced one and the same world of phenomena, i.e. the world we perceive, in several different ways: he could have made the same "outside" with several different "insides". As a consequence, the best his philosophy of nature can achieve is to correspond accurately with all the phenomena of nature, without being certain whether it accurately reflects nature's "inner" structure. All we can do is to explain the *possible*, *not actual*, *nature* of things.

Newton completely agrees with this. Both Descartes and Newton share the same premise, and both reach the same general epistemological conclusion; but despite their common ground—Newton seems to be implying—Descartes still manages to develop a natural philosophy full of atheistic potential, if not outright atheistic.

The second reason for Newton's engagement with God's creation of matter is very general: discussion of what God can do with his absolute power, how his power is related to his will, or, simply, how God created the world, was a constitutive part of natural philosophy from the Middle Ages deep into the early modern period.[25] Newton discusses God's creation of the world on several occasions, and does it for the most time without any explicit Cartesian context. This is also the case, for instance, in the famous and revealing "Query 31" of *Opticks*, the draft of which contains the above cited unsettling quote that begins with "[t]he business of Experimental

[25] See, for example Oakley 1984 and Funkenstein 1986.

Philosophy" (Newton 1675, fol. 242^{v}).[26] "All material Things", all bodies, writes Newton, seem to have been composed of the hard and solid particles variously associated "in the first Creation by the Counsel of an intelligent Agent" (Newton 1730, 402)[27] Then he goes on to claim that the world is God's sensorium, and compares the power God has to "form and reform the parts of the universe" to the power "of our own will to move the parts of our own bodies" (Newton 1730, 403).

> And since Space is divisible in infinitum, and Matter is not necessarily in all places, it may be also allow'd that God is able to create Particles of Matter of several Sizes and Figures, and in several Proportions to Space, and perhaps of different Densities and Forces, and thereby to vary the Laws of Nature, and make Worlds of several sorts in several Parts of the Universe. At least, I see nothing of Contradiction in all this. (Newton 1730, 403-404)

The next paragraph in the published text—replacing the draft version: "The business of Experimental Philosophy is to find out by experience & observation not how things were created but what is the present frame of Nature. This enquiry should proceed first by Analysis" (Newton 1675, fol. 242^{v})—continues: "As in Mathematicks, so in Natural Philosophy, the Investigation of difficult Things by the Method of Analysis, ought ever to precede the Method of Composition. This Analysis consists in making Experiments and Observations" (Newton 1675, fol. 242^{v}).

In Newton's mind, there is clearly no contradiction in doing both, i.e. figuring out "how things were created" and examining the present frame of Nature with the help of experiments and observations.

But there is more. The examination of "the present frame of nature", or of "the world as it is", would not be of any help in the enterprise of defining the body. The consequences of Newton's "totalitarian conception of God's will",[28] to which he adhered throughout his career, are such that make examination of the actual world, as far as the definition of the body is concerned, completely irrelevant. The quote from "General Scholium" cited at the beginning of this article, which affirms that we cannot know the "innermost substances" of bodies, becomes much more understandable if we take a look at its draft versions, and these, in turn, only make sense if Newton's theological voluntarism is taken into account. Newton made five (or six, if we also take the published text into consideration) slightly different versions of this formulation: all we know about the properties of

[26] See also Ducheyne 2012, 249.
[27] See also Newton 1730, 400.
[28] I am borrowing this characterization from Dempsey 2014, 93, n. 20.

things comes from the phenomena of things, and the warning: "we ought not rashly to assert that which cannot be inferred from the phenomena" (Newton 2009, fol. 361[r]; English translation from Hall and Boas Hall 1962, 360).[29] The clearest version comes from Draft C:

> From phenomena we know the properties of things, and from the properties we infer that the things themselves exist and we call them substances: but we do not have any idea of substances. We see but the shapes and colours of bodies, we hear but sounds, we touch but external surfaces, we smell odours and taste flavours; but we know the substances or essences themselves by no sense, by no reflex action, and therefore we have no more idea of them than a blind man has of colours. And when it is said that we have an idea of God or an idea of body, nothing other is to be understood than that we have an idea of the properties or attributes of God, or an idea of the properties by which bodies are distinguished from God or from each other. Whence it is that we nowhere argue about the ideas of substances apart from properties, and deduce no conclusions from the same. (Hall and Boas Hall 1962, 360–361)[30]

We are stuck with the phenomenal, not the substantial or essential level of bodies, the "surfaces" perceived by the senses, "outside" of the world, in Descartes's words, but we cannot reach the "inside" of the world (= the "things" that cannot be perceived by the senses); and, therefore, cannot come to know the true nature of things. But why is that so? Here Newton's totalitarian theological voluntarism comes in play. It is, as we have seen, because God "could create beings similar to bodies which display all their actions and exhibit all their phenomena, and yet would not be bodies in essential and metaphysical constitution" (Newton 2014, 41), because he is able to "form and reform the parts of the universe," and because he is also able to "vary the Laws of Nature, and make Worlds of several sorts in several Parts of the Universe," and because "[h]e is what he is by the necessity of nature, all other things are what they are by the power of his will" (Newton 2002).

Conclusion

One of the most challenging aspects of Newton's genesis, a thought experiment on how God could have created matter/bodies, intimately interwoven with theological voluntarism, are its epistemological consequences. John Henry thus affirms that

[29] For the original Latin, also see Hall and Boas Hall 1962, 356.
[30] For the original Latin, see Hall and Boas Hall 1962, 356-357.

> the voluntarist emphasis upon God's freedom of operation is associated with a belief in the radical contingency of the natural world and the concomitant belief that we can only understand God's creation *a posteriori*, by examining it and drawing empirically-based conclusions as to what he actually did, or as to what kind of world he created. (Henry 2009, 81)

This seems to be a fairly accurate characterization of Newton's project of experimental philosophy—but, unfortunately, it neglects one very important epistemological dimension. As we have seen, Newton's omnipresent and very active God could have created/formed a world of beings—and is always able to recreate/reform it, including its natural laws—, which are completely like bodies in all their aspects and actions but are not essentially bodies. This is the world of phenomena. And we only have access to the world of phenomena, i.e. the world we can perceive. Our perception, our senses, cannot reach beyond the phenomena, we cannot reach the inner nature of things. Newton's genesis thought experiment in *De gravitatione* and other texts, which develop identical ideas in different terms, or represent variations of similar themes, show not only "abiding skepticism about the possibility of ever finding out the nature of body, or corporeal substance," as claims Gabbey, but, I would argue, positively preclude any such possibility. Bodies are impossible to define, or, more precisely, they can be defined only conditionally: "*If* [these created beings] are bodies, *then* we can define bodies as determined quantities of extension which omnipresent God endows with certain conditions" (Gabbey 2011)—, but we can be never sure that they really are. That is why even when Newton does define body, as, for example, in the draft to the third edition of *Principia*, he does not claim the definition to be metaphysically true, but only "commonly received": "Definition II. I call a body every mobile and tangible thing that is resisted by things touching it, and whose resistance, if it be large enough, can be felt. It is in this sense that the name body has always been commonly received" (McGuire 1995, 115). The same also goes on in *De gravitatione*. In addition to the above cited conditional definition of bodies, the Definition 2 (at the beginning of *De gravitatione*) has it: "Body is that which fills place" (Newton 2014, 27). But Newton quickly explains that his definition regards body "not in so far as it is a physical substance endowed with sensible qualities, but only in so far as it is extended, mobile, and impenetrable" (Newton 2014, 27), i.e. he did not define body "in a philosophical manner, but abstracting the sensible qualities /…/, I have postulated only the properties required for local motion" (Newton 2014, 27). Here again, we have a definition of body, but not a definition of its nature or essence, but a definition which is appropriate for the consideration of the local motion of bodies only.

So, theological voluntarism, at least in Newton's case, puts some severe restrictions on our knowledge of the world. How then to proceed? Here a comparison between Descartes and Newton is in order. As we have seen, Descartes and Newton share, at least on the fundamental level, the same premise: Descartes' and Newton's God was free to have made the same "outside"/phenomena of the world with several different "insides", i.e. metaphysical, of the world, or, as Gabbey puts it in reference to Newton: "God was free to have created a quite different meta-phenomenal world corresponding to exactly the same phenomena" (Gabbey 2011, 437). In Descartes' case, this means that some of his explanations are "at least morally certain,"[31] even if they may be uncertain in relation to the absolute power of God, but there are also some matters, "even in relation to the things in nature" (Descartes 2009, Part IV, art. 206, 290), which he considers more than morally certain, i.e. they are absolutely certain. Absolute certainty arises "when we believe that it is wholly impossible that something should be otherwise than we judge it to be" (Descartes 2009, Part IV, art. 206, 290), and is based on metaphysical foundations: the supreme goodness of God; that God is not a deceiver and hence the faculty he gave us to distinguish truth and falsehood cannot lead us into error if used properly, i.e. when we perceive something distinctly. Such are mathematical demonstrations, knowledge that material things exist, and all evident reasoning about material things, and perhaps even Descartes' more specific natural philosophical results, given that "they have been deduced in an unbroken chain from the first and simplest principles of human knowledge" (Descartes 2009, Part IV, art. 206, 290).

Although Descartes admits that God could have created the same world of phenomena with different "insides", this epistemological problem is overturned with emphasis on God's goodness and what follows therefrom. This is the crucial distinction between Newton and Descartes. Newton is a totalitarian voluntarist; his God can do whatever he wishes short of creating a logical contradiction. Descartes is an intellectualist: his God acts according to his goodness. Descartes' God cannot deceive us as far as our perceptual capabilities are concerned (when we have clear and distinct perception), while Newton's can deceive us. Newton is therefore "reluctant to say *positive* what the nature of bodies is," while *positive* captures both meanings Gabbey discusses. Newton "is unwilling to give"—read "cannot give"—"any account of the nature of bodies that is not in terms of real and true essences arising from real causes and confirmed by sensory evidence" (Gabbey 2011, 429, n. 15), and he at the same time declares "his unwillingness to indulge in arbitrary theorizing on the nature of bodies"

[31] Moral certainty is certainty that is sufficient for application in ordinary life.

(Gabbey 2011, 429, n. 15), or, in other words, "he is unwilling to feign hypotheses on the question" (Gabbey 2011, 429, n. 15). Here is the ultimate reason for Newton's refusal of Descartes' hypothetical natural philosophy and his adoption of an experimental one—but that is another story.

Bibliography

Ariew, Roger. 2005. "Descartes's Fable and Scientific Methodology". *Archives internationales d'histoire des sciences*, Vol. 55, No. 154: 127-138.

Ariew, Roger. 2010. "The New Matter Theory and Its Epistemology Descartes (and Late Scholastics) on Hypotheses and Moral Certainty". In *Vanishing Matter and the Laws of Motion*, edited by Peter Anstey and Dana Jalobeanu, 31-46. New York: Routledge.

Bennett, Jonathan, and Remnant, Peter. 1978. "How Matter Might at First Be Made". *Canadian Journal of Philosophy* (Supplementary Vol. 4): 1-11.

Biener, Zvi, and Smeenk, Christopher. 2012. "Cotes's Queries: Newton's Empiricism and Conceptions of Matter". In *Interpreting Newton*, edited by Andrew Janiak and Eric Schliesser, 103-137. Cambridge: Cambridge University Press.

Birch, Thomas. 1757. *The History of the Royal Society*, Vol. 3. London: For A. Millar in the Strand, 247-305. Available at: http://www.newtonproject.ox.ac.uk/view/texts/normalized/NATP00002.

Brading, Katherine. 2012. "Newton's Law-Constitutive Approach to Bodies: A Response to Descartes". In *Interpreting Newton*, edited by Andrew Janiak and Eric Schliesser, 13-32. Cambridge: Cambridge University Press.

Craig, John. 1727. *Letter from John Craig to unidentified recipient* (Keynes Ms. 132). King's College, Cambridge. Available at: http://www.newtonproject.ox.ac.uk/catalogue/record/THEM00030.

Curley, Edwin. 1993. "Certainty: Psychological, Moral, and Metaphysical". In *Essays on the Philosophy and Science of René Descartes*, edited by Stephen Voss, 11-30. Oxford: Oxford University Press.

Damerow, Peter. 2016. "Space and Matter in Early Modern Science: The Impenetrability of Matter". In *Spatial Thinking and External Representation: Towards a Historical Epistemology of Space*, edited by Matthias Schemmel, 175-190. Berlin: Max Planck Institute for the History of Science.

Davis, Edward B. 1991. "Newton's Rejection of the 'Newtonian World View': The Role of Divine Will in Newton's Natural Philosophy". *Science and Christian Belief*, Vol. 3, No. 1: 103-118.

Dempsey, Liam P. 2006. "Written in the flesh: Isaac Newton on the mind-body relation". *Studies in History and Philosophy of Science Part A*, Vol. 37, No. 3: 420-441.

Dempsey, Liam P. 2009. "Thinking-Matter Then and Now: The Evolution of Mind-Body Dualism". *History of Philosophy Quarterly*, Vol. 26, No. 1: 43-61.

Dempsey, Liam P. 2014. "Newtonian Idealism: Matter, Perception, and the Divine Will". *The Southern Journal of Philosophy*, Vol. 52, No. 1: 86-111.

Descartes, Rene. 2009. *Principles of Philosophy*. In *The Philosophical Writings of Descartes*, Vol. 1. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch, 177-292. Cambridge: Cambridge University Press.

DiSalle, Robert. 2016. "Newton's Philosophical Analysis of Space and Time". In *The Cambridge Companion to Newton* (2nd edition), edited by Rob Iliffe and George E. Smith, 34-60. Cambridge: Cambridge University Press.

Dobbs, Betty J. T. 1991. The Janus Faces of Genius: The Role of Alchemy in Newton's Thought. Cambridge: Cambridge University Press.

Ducheyne, Steffen. 2012. *The Main Business of Philosophy*. Dordrecht: Springer.

Force, James E., and Hutton, Sarah. 2004. *Newton and Newtonianism*. Dordrecht: Kluwer.

Funkenstein, Amos. 1986. Theology and the Scientific Imagination from the Middle Ages to the Seventeenth Century. Princeton: Princeton University Press.

Gabbey, Alan. 2011. "The Term *materia* in Newton and in the Newtonian Tradition". In *Materia: Atti del XIII colloquio internazionale del lessico intellettuale Europeo*, edited by Delfina Giovannozzi and Marco Veneziani, 423-445. Florence: Leo S. Olschki.

Garber, Daniel; Henry, John; Joy, Lynn; and Gabbey, Allan. 1998. "New Doctrines of Body and Its Powers, Place and Space". In *The Cambridge History of Seventeenth Century Philosophy* (Vol. 1), edited by Daniel Garber and Michael Ayers, 553-623. Cambridge: Cambridge University Press.

Gorham, Geoffrey. 2011a. "How Newton Solved the Mind-Body Problem". *History of Philosophy Quarterly*, Vol. 28, No. 1: 21-44.

Gorham, Geoffrey. 2011b. "Newton on God's Relation to Space and Time: The Cartesian Framework". *Archiv für Geschichte der Philosophie*, Vol. 93, No. 3: 281-320.

Guerlac, Henry. 1983. "Theological Voluntarism and Biological Analogies in Newton's Physical Thought". *Journal of the History of Ideas*, Vol. 44, No. 2: 219-229.

Hall, Rupert A. and Boas Hall, Marie (eds.). 1962. Unpublished scientific papers of Isaac Newton: a selection from the Portsmouth collection in the University Library, Cambridge. Cambridge: Cambridge University Press.

Harrison, Peter. 2002. "Voluntarism and Early Modern Science". *History of Science*, Vol. 40, No. 1: 63-89.

Harrison, Peter. 2004. "Was Newton a Voluntarist?". In *Newton and Newtonianism: New Studies*, edited by J. E. Force and S. Hutton, 39-63. Dordrecht: Kluwer.

Harrison, Peter. 2009. "Voluntarism and the Origins of Modern Science: A Reply to John Henry". *History of Science*, Vol. 47, No. 2: 223-231.

Harper, William. 2016. "Newton's Argument for Universal Gravitation". In *The Cambridge Companion to Newton*, edited by Rob Iliffe and George E. Smith, 229-260. Cambridge: Cambridge University Press.

Heimann, Peter M. 1978. "Voluntarism and Immanence: Conceptions of Nature in Eighteenth-Century Thought". *Journal of the History of Ideas*, Vol. 39, No. 2, 271-283.

Henry, John. 2009. "Voluntarist Theology at the Origins of Modern Science: A Response to Peter Harrison". *History of Science*, Vol. 47, No. 1: pp. 79-113.

Henry, John. 2013. "The Reception of Cartesianism". In *The Oxford Handbook of British Philosophy in the Seventeenth Century*, edited by Peter R. Anstey, 116-143. Oxford: Oxford University Press.

Henry, John. 2014. "Newton and Action at a Distance between Bodies: A Response to Andrew Janiak's 'Three Concepts of Cause in Newton'". *Studies in History and Philosophy of Science Part A*, Vol. 47: 91-97.

Iliffe, Rob. 1995. "'That Puzleing Problem': Isaac Newton and the Political Physiology of Self'. *Medical History*, Vol. 39, No. 4: 433-458.

Iliffe, Rob. 2017. Priest of Nature: The Religious Worlds of Isaac Newton. Oxford: Oxford University Press.

Jalobeanu, Dana. 2007. "Space, Bodies and Geometry: Some Sources of Newton's Metaphysics". *Zeitsprünge: Forschungen zur frühen Neuzeit*, Vol. 11, No. 1-2: 81-113.

Jalobeanu, Dana. 2011. "The Cartesians of the Royal Society: The Debate over Collisions and the Nature of Body (1668-1670)". In *Vanishing Matter and the Laws of Nature*: *Descartes and Beyond*, edited by Dana Jalobeanu and Peter Anstey, 103-129. London: Routledge.

Jalobeanu, Dana. 2013. "The Nature of Body". In *The Oxford Handbook of British Philosophy in the Seventeenth Century*, edited by Peter Anstey, 213-239. Oxford: Oxford University Press.

Janiak, Andrew. 2008. *Newton as Philosopher*. Cambridge: Cambridge University Press.

Kochiras, Hylarie. 2011. "Gravity's Cause and Substance Counting: Contextualizing the Problems". *Studies in History and Philosophy of Science Part A*, Vol. 42, No. 1: 167-184.

Kochiras, Hylarie. 2013. "By ye divine arm: God and Substance in De gravitatione". *Religious Studies*, Vol. 49, No. 3: 327-356.

Laymon, Ronald. 1978. "Newton's Bucket Experiment". *Journal of the History of Philosophy*, Vol. 16, No. 4: 399-413.

Locke, John. 1735. *Essai philosophique concernant l'entendement humain* (Vol. 2). Translated by Pierre Coste. Amsterdam: Pierre Mortier.

McAllister, James. 1996. "The Evidential Significance of Thought Experiments in Science". *Studies in History and Philosophy of Science*, Vol. 27, No. 2: 233-250.

McAllister, James. 2005. "The Virtual Laboratory: Thought Experiments in Seventeenth-Century Mechanics". In *Collection - Laboratory - Theater: Scenes of Knowledge in the 17th Century*, edited by Helmar Schramm, Ludger Schwarte, and Jan Lazardzig, 35-56. New York: Walter de Gruyter.

McAllister, James. 2013. "Thought Experiment and the Exercise of Imagination in Science". In *Thought Experiments in Philosophy, Science, and the Arts*, edited by Mélanie Frappier, Letitia Meynell, and James Robert Brown, 11-29. London: Routledge.

McGuire, James E. 1982. "Space, Infinity, and Indivisibility: Newton on the Creation of Matter". In *Contemporary Newtonian Research*, edited by Zev Bechler, 145-190. Dordrecht: Reidel.

McGuire, James E., and Tamny, Martin, eds. 1983. *Certain Philosophical Questions: Newton's Trinity Notebook*. Cambridge: Cambridge University Press.

McGuire, James E. 1995. "Body and Void and Newton's *De Mundi Systemate*: Some New Sources". In *Tradition and Innovation: Newton's Metaphysics of Nature*, 103-150. Dordrecht: Kluwer. First published in 1966, *Archive for History of Exact Sciences*, Vol. 3, No. 3: 206-248.

Newton, Isaac. 1661-65. *Trinity College Notebook* (Ms. Add. 3996). Cambridge: Cambridge University Library. Available at: https://cudl.lib.cam.ac.uk/view/MS-ADD-03996/1.

Newton, Isaac. 1675. *Draft Versions of 'The Queries' (*Ms. Add. 3970). Cambridge University Library. Available at: http://www.newtonproject.ox.ac.uk/view/texts/normalized/NATP00123.

Newton, Isaac. 1687/1999. *The Principia: Mathematical Principles of Natural Philosophy*. Translated by Bernard Cohen and Anne Whitman. Berkeley: University of California Press.

Newton, Isaac. 1730. *Opticks: or, A treatise of the reflexions, refractions, inflexions and colours of light*. London: Printed for William Innys at the West-End of St. Paul's.

Newton, Isaac. 1962. "De gravitatione et aequipondio fluidorum". In *Unpublished scientific papers of Isaac Newton: a selection from the Portsmouth collection in the University Library, Cambridge*, edited by A. Rupert Hall and Marie Boas Hall, 89-121 (Latin text) and 121-156 (English translation). Cambridge: Cambridge University Press.

Newton, Isaac. 2002. *Irenicum, or Ecclesiastical Polyty tending to Peace* (Keynes Ms. 3). Cambridge: King's College. Available at: http://www.newtonproject.ox.ac.uk/view/texts/normalized/THEM00003.

Newton, Isaac. 2006. "Dibner Ms. 1031 B". At *The Chymistry of Isaac Newton*, edited by William R. Newman. Available at: http://purl.dlib.indiana.edu/iudl/newton/ALCH00081.

Newton, Isaac. 2009. *Draft C of the 'Scholium Gnerale'* (Ms. Add. 3965.12, ff. 361-362). Cambridge: Cambridge University Library. Available at: http://www.newtonproject.ox.ac.uk/view/texts/normalized/NATP00059.

Newton, Isaac. 2014. *Philosophical Writings* (2nd edition), edited by Andrew Janiak, Cambridge: Cambridge University Press.

Oakley, Francis. 1961. "Christian Theology and the Newtonian Science: The Rise of the Concept of the Laws of Nature". *Church History*, Vol. 30, No. 4: 433-457.

Oakley, Francis. 1984. Omnipotence, Covenant, and Order: An Excursion in the History of Ideas from Abelard to Leibniz. Ithaca: Cornell University Press.

Osler, Margaret. 1994. Divine Will and the Mechanical Philosophy: Gassendi and Descartes on Contingency and Necessity in the Created World. Cambridge: Cambridge University Press, 118-170.

Palmerino, Carla Rita. 2013. "The Composition of Space, Time and Matter According to Isaac Newton and John Keill". In *The Mechanization of Natural Philosophy*, edited by Daniel Garber and Sophie Roux, 117-142. Dordrecht: Springer.

Pourciau, Bruce 2016. "Instantaneous Impulse and Continuous Force: The Foundations of Newton's Principia". In *The Cambridge Companion to Newton*, edited by Rob Iliffe and George E. Smith, 93-186. Cambridge: Cambridge University Press.

Roux, Sophie. 2011. "Introduction: The Emergence of the Notion of Thought Experiments". In *Thought Experiments in Methodological and*

Historical Contexts, edited by Katerina Ierodiakonou and Sophie Roux, 1-33. Leiden and Boston: Brill.

Schliesser, Eric. 2010. "Without God: Gravity as a Relational Property of Matter in Newton". In *Vanishing Matter and the Laws of Nature*, edited by Dana Jalobeanu and Peter R. Anstey, 80-100. London: Routledge.

Slowik, Edward. 2009. "Newton's Metaphysics of Space: A 'Tertium Quid' betwixt Substantivalism and Relationism, or merely A 'God of the (Rational Mechanical) Gaps'?". *Perspectives on Science*, Vol. 17, No. 4: 429-456.

Stein, Howard. 2016. "Newton's Metaphysics". In *The Cambridge Companion to Newton*, edited by Rob Iliffe and George E. Smith, 321-281. Cambridge: Cambridge University Press.

Tamny, Martin. 1979. "Newton, Creation, and Perception". In *Isis*, Vol. 70, No. 1: 48-58.

PART 2:

PHILOSOPHICAL ARGUMENTS, IMAGINATION AND THOUGHT

CHAPTER NINE

PRECURSORS OF THE ANTHROPIC REASONING IN LATE PYTHAGOREANISM AND ANCIENT ATOMISM

MILAN M. ĆIRKOVIĆ

> Is there a thing of which it is said, 'See, this is new'?
> It has been already, in the ages before us.
> There is no remembrance of former things,
> nor will there be any remembrance of later things yet to happen
> among those who come after.
> Ecclesiastes 1:10-11

In recent decades, there has been a surge of interest in anthropic reasoning, motivated by several similar but independent developments in contemporary science and philosophy. In particular, the awareness that we are likely living in a much large cosmological whole—the *multiverse*—has increased as a consequence of success of the cosmological inflationary paradigm in explaining puzzles related to the initial conditions of our observable domain (Ellis, Kirchner, and Stoeger 2004; Vilenkin 2006; Carr 2007; Kragh and Longair 2019). Also, the advent of new radical disciplines such as quantum cosmology, quantum information theory, and quantum computing, has boosted the adherence to the many-worlds ontologies. This development is finely summarized in a popular review by Tegmark (2003), which offers a tentative classification scheme for easier thinking about the grand cosmological ensembles. The Level I multiverse would be an extension of our universe beyond the cosmological horizon(s), something which is uncontroversial even in the standard Friedmann models. The Level II multiverse would encompass all cosmological domains ("universes") originating through cosmological inflation, or a similar symmetry-breaking process. On a deeper level still, if many-worlds interpretations/theories of quantum mechanics—most notably, Everett's "no collapse" theory—are

valid, the totality of all branches of the wavefunction of the universe would constitute the Level III multiverse. Finally, any structure *larger than that* would belong on the Level IV—and in particular this applies to Tegmark's own "ultimate mathematical ensemble" theory (Tegmark 1998, 2008) or the philosophical doctrine of modal realism (Lewis 1986). Note that each level contains lower levels as special cases in which at least one global symmetry is broken. Therefore, as is often the case with the spontaneous symmetry breaking, *a part is more complex than the whole*, with multiple philosophical consequences which have not been elaborated to this day.

Multiverse theories are often criticized for their "overwhelming ontologies"; it has also been suggested that they violate some unspoken and deep rules about "how we should be doing science". This misleading and self-centred prejudice should be confronted and dispersed—and it is exactly the history of philosophy and the history of (early) science which could help in the task. In contrast to the views of the critics, modern analysis of observation-selection effects, also known as the anthropic reasoning, is not some new-fangled philosophical fashion without valid precedent. In fact, quite the opposite is the case; anthropic reasoning is quite old and it traces back to classical antiquity, although in often disguised and intrinsically complex forms. In the rest of this contribution, I discuss several of the instances of what could reasonably be dubbed *ancient anthropic reasoning*. This is in agreement with both abovementioned increase of interest in multiverse theories, and the renewed interest for ancient science, in particular ancient precursors to modern physical cosmology and fundamental physics (e.g. Gregory 2007; Graham 2013; Nicolaides 2020). In contrast to the narrow, constraining, and nihilistic views of postmodernists and social constructionists, which have until recently dominated the history of science and technology and revelled in proclaiming any different views "Whiggish history", researchers of the 21st century feel free of these postmodernist shackles and are ready to explore and trace origins of the very real progress of science. As Andrew Gregory emphasizes:

> [T]here are perennial philosophical and scientific problems relating to cosmogony. This is not to say that these problems are insoluble, or that no progress has been made in relation to them. On the contrary, I believe we now have a much more sophisticated understanding of these problems and a greater and more sophisticated array of possible solution, even if we often lack definitive solutions… While there may be perennial core problems in cosmogony, they manifest themselves in varying forms at different times and in different circumstances. The problems as perceived by the ancient Greeks are related to but not identical to the problems addressed by modern cosmology. (Gregory 2007, 2)

The most obvious of these cosmological problems that deals, explicitly or implicitly, with the properties of observers is the problem of infinite past. Today's standard cosmological model (Weinberg 1972 and 2008) contains a sharp asymmetry between spatial and temporal infinity. Almost all cosmologists agree that the universe began in a singular—or "nearly singular" (intuitive understanding will need to suffice for now)—state about 14 billion years ago, popularly called the Big Bang. The universe has expanded, as we now know, in an accelerated manner for the last about 4 billion years, and it will continue to expand indefinitely. So, there is no past temporal infinity, while the future temporal infinity is endorsed in the standard cosmology. On the other hand, there is absolutely no strictures against spatial infinity—while in the new standard ΛCDM model spatial slices could be either finite or infinite, most cosmologists have inherited the old view, characterizing classical Friedmann models and favouring open/flat universe topologies, which contain spatial infinity. That this has some strange philosophical consequences, like the infinity of exact doubles of you and me, has been known for quite some time (e.g. Ellis and Brundrit 1979), but has not been traditionally considered too problematic. After all, the existence of cosmological horizons, as a rather generic feature of relativistic cosmological models, precludes this *potential* strangeness becoming actual; or at least it postpones the problem into an excessively distant future.

However, the asymmetry between spatial and temporal infinities is an inheritance of ancient Greek thinkers. Although we are lacking exact data, Pythagoras and the early Pythagoreans supposedly argued for spatial infinity of the world, although this has become manifest only much later, in the thought of Philolaus and Archytas about 400 BC. In particular, Archytas's thought experiment with "spear" or "staff" extended at the boundary of the spatial finite universe has been extremely influential in widespread acceptance of the infinity of space (Ierodiakonou 2011). This was, of course, in complete accordance with the concept of infinite (spatially-extended) worlds of Anaximander and Democritus, as well as subsequent atomists and even Stoics. Note that this conclusion by-passes the contentious questions of what does constitute a "world" or how one could talk about vacuum or empty space—it is enough that the concept has spatial extension whatsoever. All this has been inherited in modern physical cosmology, in spite of the fact that Einstein was the first to essentially circumvent Archytas's *Gedankenexperiment* by showing, in his first 1917 cosmological model, that the curved space of general relativity allows for a finite space without edges or boundaries. An interesting conjecture for further research in history of ideas would be to show to what extent does the

readiness of modern physical cosmology to accept spatial infinity even within a single cosmological domain (a "universe") stem from this ancient philosophical tradition.

In contrast, there are all reasons to believe that the philosophers of antiquity were keenly aware of difficult problems with the (past) *temporal* infinity from a very early epoch on. Part of this has been an intuitive understanding that the infinite past is antithetical to evolutionary worldviews; the latter existed in the classical antiquity since Anaximander and his famous fragment A30, which seems to argue that humans originated from marine animals. Subsequent elaborated evolutionary schemes are to be found in Empedocles's and Anaxagoras' doctrines. These are p*rima facie* incompatible with naïve views of the past temporal infinity of the form: the physical world has always existed in a more-or-less similar form as it exists now. This was, in the context of the steady-state theory, dubbed the "perfect cosmological principle" (Balashov 1994; Kragh 1996). In such a universe, it would be impossible to speak of the beginning of any temporal series or of history of anything in strict sense; it would be at least implausible to have an event or an item described as "the oldest" in anything but the vernacular. There are ways of going around this—as Empedocles did with his cyclic cosmology, the first of its kind in the history of ideas, or the "world conflagration" (ἐκπύρωσις) of the Stoics (O'Brien 1969; Rosenmeyer 1989). These ways are, however, not at all obvious. On the contrary, they can lead to strange conclusions, which are illustrated in many ancient sources. Consider the very first passage of Thucydides's *History of the Peloponnesian War*:

> Thucydides, an Athenian, wrote the war of the Peloponnesians and the Athenians as they warred against each other, beginning to write as soon as the war was on foot, with expectation it should prove a great one and most worthy the relation of all that had been before it; conjecturing so much both from this, that they flourished on both sides in all manner of provision, and also because he saw the rest of Greece siding with the one or the other faction, some then presently and some intending so to do. For this was certainly the greatest commotion that ever happened among the Grecians, reaching also to part of the barbarians and, as a man may say, to most nations. For the actions that preceded this and those again that are yet more ancient, though the truth of them through length of time cannot by any means clearly be discovered, yet for any argument that, looking into times far past, I have yet light on to persuade me, I do not think they have been very great, either for matter of war or otherwise. (*History of the Peloponnesian War*, 1.1)

Essentially, the great historian here defends—from our modern point of view a bizarre and ludicrous—thesis that before his time, i.e. around 450 BC, *nothing of importance has happened in history*. Of course, he does appeal to the empirical difficulties of finding out the truth about distant past, which "cannot by any means clearly be discovered", but still this prejudice can hardly fail to surprise a modern reader. Oswald Spengler called this statement "outrageous" and used it to argue for non-scientific, and essentially mythological character of ancient historiography. It might, however, also be interpreted as a way of dealing with the tension between the assumed past temporal infinity and the obvious fact of *finitude* of any human history or narrative. An enlightened and rational observer, such as Thucydides, could not fail to notice how outright impossible it is to bring in accord the following theses:

(i) the world had always existed in more-or-less similar form as it was about 450 BC; and
(ii) there had been just a few notable events (the Trojan War, the Persian invasion, etc.) before about 450 BC.

The solution he hints at in the continuation of Book 1 is that before the Trojan and Theban War, all humans lived in the state of "original barbarism", without civilization and organized society, and therefore had nothing of significance to report. Of course, he might have also been aware of the difficulties with such a hypothesis, so it was wise not to go into any detail; for instance, why did the civilization start only at that particular assumed epoch (about or shortly before the Theban War) and not in the infinite number of centuries prior? In the temporal infinity, any starting point is as good as any other.

The same uncertainty motivated Lucretius (and, supposedly, Epicureans and other ancient atomists), to advance a clear-cut "short history" argument for the finite past, one of the very first arguments based on observation selection effects. In Book V of *De Rerum Natura*, he wrote the following intriguing verses:

If there had been no origin-in-birth
Of lands and sky, and they had ever been
The everlasting, why, ere Theban war
And obsequies of Troy, have other bards
Not also chanted other high affairs?
Whither have sunk so oft so many deeds
Of heroes? Why do those deeds live no more,
Ingrafted in eternal monuments

Of glory? Verily, I guess, because
The Sun is new, and of a recent date
The nature of our universe, and had
Not long ago its own exordium. (Lucretius 1997)

For the highly scientific-minded Lucretius, the shortness of human history alluded to by Thucydides *is* indeed very strange in the face of the conventional assumption of past temporal infinity and the eternal existence of the world. It does require an explanation! Although the references to "eternal monuments" and "other bards" may sound naive to the modern-day reader, it is clear that Lucretius had in mind any form of *transmission of information* from the past to the present. To a Roman and a poet, monuments and ancient epics have been such "time capsules", reaching the present from the past. There is no substantial difference between those and our own ideas of transferring information through time and space via artefacts (Jarvis 2003; Rahman 2009). Information-carrying artefacts have been suggested as a method of communication with extra-terrestrial intelligence as well (Rose and Wright 2004); extremely long durability of information-carriers has been recently considered in the context of space engineering (Guzman, Hein, and Welch 2017).[1]

Any such method of transfer, no matter how inefficient, would tend to transfer an infinite amount of information from an infinite past. Lucretius's empirical assessment of the surrounding world clearly shows the absence of such information. On the contrary, even the oldest monuments—the Pyramids of Egypt, for example—are of negligible age in comparison to the completely imaginable millions or billions of years, not to mention temporal *infinity*. (One could reasonably suppose that the famous tract of Archimedes on large numbers, *The Sand Reckoner*, was well-known to Lucretius.) Therefore, an explanation is needed. Could they all be destroyed? All infinite number of them? A pyramid, say, twenty times larger than the one of Khufu?[2] The simplest explanation, as Lucretius was highly aware, is to treat the argument as a *reductio ad absurdum* of the starting hypothesis (eternal nature of the world) and to assume that the world—and, hence, all

[1] These points are worth keeping in mind, since the questions are very similar to some posed in contemporary astrobiology and the Search for Extra-Terrestrial Intelligence (SETI) studies, as argued by the author in Ćirković (2012). Even if we accept finite past of our universe, its age and the ages of relevant structures like galaxies and stars are so immensely larger than the timescale of human history, that for many practical purposes we may treat them as *effectively infinite*.

[2] Such a building would still be very far from the outer envelope of what is achievable with even relatively conservative engineering, a question which bothered Milutin Milanković, among others (see Milanković 1956).

of its features, including humans and their cultures—is of finite age, and a rather short one.

Note two important aspects of this argument of Lucretius: (I) there is a necessity in treating the present epoch as special, at least to a degree. This is implicit: we know that present-day humans (or at least those humans *sufficiently like their present-day analogues*) build monuments, wage wars, and have bards singing about them—so we use exactly these criteria to compare them with other, hypothetical, temporal slices. Is that not a wee bit problematic? By analogy, some critics of the modern SETI programs have argued that we are intentionally searching only for those aliens that are sufficiently similar to ourselves; for instance, those who build radio-telescopes and exchange meaningful messages using radio waves (for this kind of criticism, see e.g. Rescher 1985; Diamond 1992). The criticism is perhaps too strident for its own good. Lucretius certainly could not conceive of many particular realizations of his general category of monuments or of his category of bardic chants, which need not be necessarily thought of as *Ozymandias*-like broken sculpture or an *Iliad*-like epic poem; this contingent fact does not entail that, upon encountering ruins of a nuclear power station or a dense sheet of atonal music, he would not recognize them as artefacts. Of course, if we consider gradual increase in "strangeness" of an artefact, at one point it might cease to be recognizable as such by any human observer. This suggests that temporal Copernicanism, a tendency to regard our own cosmological epoch as typical, is of limited value in general philosophical considerations. A number of "eternal monuments" could, if strange enough, pass unnoticed to Lucretius's (or indeed our own) eyes (see Hoyle 1983). For more detailed arguments against temporal Copernicanism, see Ćirković and Balbi (2020) and references therein.

(II) Explicit mention of the origin of the Sun testifies that Lucretius rejected the Aristotelian (and subsequently Scholastic) view about immutability of "translunar spheres", which formed the backbone of the classical Aristotelian-Ptolemaic geocentric model of the universe. One might downplay this fact since atomists, including Lucretius, clashed with Aristotle on many other, allegedly bigger, issues. This is important, however, for subsequent attempts to "limit the damage" at various stages of the perennial struggle between the evolutionist and the dogmatic metaphysical thinking. Now, since Epicureans allowed for *objectively random* (stochastic) events in the world, we are entitled to speculate to what extent our position in the world is indeed a random state of affairs. In any case, if the Sun is relatively new, we may interpret it as either a consequence of the finite past as such, or it is only our local environment (for instance the solar system), which is of recent origin. The distinction is absolutely crucial for modern physical

cosmologies for several reasons, one being that some of the alternatives to the standard relativistic cosmology included *local* discontinuities or even local singularities (e.g. Hoyle, Burbidge, and Narlikar 1993, 1994). Although such schemes do not pass muster today, we should be wary about the premature formation of an orthodoxy and the extent to which we could expect new alternatives to the standard physical cosmology to emerge in the future (Kragh 1996, 2007; Ćirković and Perović 2018 and references therein).

A *global* alternative has been offered by Empedocles. He was the first to propose a way to restore agreement between past temporal infinity (at least in a formal sense) and the finite age of everything we locally observe. Consider: what exactly constitutes the past temporal infinity from a physicalist point of view? Not any perishable structures such as our bodies, or even the Lucretian "eternal monuments". On the contrary, it must be the most fundamental constituents of matter. For Empedocles, those were the four classical elements, as well as the two fundamental and opposing forces (Love and Strife, i.e. attractive and repulsive interactions). If these very basic entities have always existed and will forever exist, then we can claim, at least formally, past (and future) temporal infinity for the world. In particular, the fragment B 16 of the Diels collection reads (according to Burnet's 1892 translation), "For of a truth they [Love and Strife] were aforetime and shall be; nor ever, methinks, will boundless time be emptied of that pair" (Burnet 1892, 223) So we have

persistence of forces + "boundless time".

It is similar with material constituents—four elements (an Atomist's version could be constructed equally well with atoms and forces between them, or perhaps just a single force, switching signs depending on distance, as Roger Boscovich did much later). In this sense, the cosmology of Empedocles is *uniformitarian*.

Empedocles suggests two complementary perspectives, however. In one of them, the usual temporal becoming is a kind of illusion, an artefact of human cognizance *or even human language*. "There is no coming into being of aught that perishes, nor any end for it in baneful death; but only mingling and change of what has been mingled. Coming into being is but a name given to these by men" (Burnet 1892, 220). This is an evolutionist, and even progressive, perspective on both physical and biological universe: "For out of these have sprung all things that were and are and shall be—trees and men and women, beasts and birds and the fishes that dwell in the waters, yea, and the gods that live long lives and are exalted in honor" (Burnet 1892,

223). Note that, like in the later Epicurean philosophy, and in sharp contrast with the Platonic tradition, Empedocles's deities are fully material, physical, and indeed *naturalistic* beings. A modern reader could relate much better to this ancient concept, since both in *Purifications* and *On Nature* we encounter description of gods as, in modern terms, *posthumans* or even old and advanced extra-terrestrial intelligences.

The other perspective is cosmological: there is a global, eternal cycling constituting the history of the universe (O'Brien 1969; Brown 1984). The cyclic motion of matter is governed by the change in relative intensities of the two interactions: in one part of the cycle, Love is on the ascendance and things are gradually brought into complete unity; in the next part, Strife rises from the minimal value and increases until everything is as widely dispersed and diffused as possible. In both these singular states, complete unity and complete disunity, all previously existing things and structures are obliterated, no life is possible, and no information can survive into the subsequent epoch. So, as much as it is uniformitarian, Empedocles's cosmology is *catastrophic* as well! Recurrent singular states are indeed global catastrophes, vaguely analogous to concepts such as the Big Bang, the Big Crunch, or the recently introduced Big Rip (Caldwell, Kamionkowski, and Weinberg 2003) in contemporary physical cosmology. And, in contrast to "conflagrations" of the Stoics, there is a clear physical picture of why these singular states occur in the first place.

Empedocles's cosmology contains an essential reference to observers and physical preconditions for their existence: he states the obvious fact that there can be no life and hence no observers like humans (or even gods for that matter) in the singular states. Neither could any information about previous events pass through such states—hence no Lucretian "eternal monuments" or bardic chants of arbitrary high age can reach us. While the universe is formally eternal—in the abovementioned sense that its basic constituents exist at any point in time—the age of everything else is sharply limited by the duration of the Great Cycle.

This elegant solution to the problem of arbitrarily old information is obviously one of the highlights of ancient cosmological thought. What Empedocles describes is essentially a model of the multiverse: the set of temporally successive cosmological domains, separated by singularities in which no structure is possible and no transfer of information occurs. Singular states essentially erase information about previous the evolution and offer a kind of "blank slate", thus removing the paradoxical consequences. There is no substantial difference between Empedocles's multiverse and a set of cosmological domains co-existent in time but topologically unconnected in space. Hence, a kind of symmetry between

spatial and temporal infinity is restored. Alternatively, the Great Cycle could be regarded as a precursor to the oscillating cosmologies that have been popular in the course of the 20th century (e.g. Barrow and Dabrowski 1995). While these oscillating models encounter practical difficulties—in particular those related to entropy increasing from one cycle into another; ironically due to the very same physical processes that ensure the erasure of inter-cycle information—they still present a fine pedagogical tool for many aspects of modern cosmology and, arguably, one of the most intuitive and palatable "multiverses" available.

The most important ingredient in all of these ideas is the implied physical nature of observership, which tallies well with the materialism of early Presocratics, as well as the later Epicurean atomists. This is important in view of the general physicalist construals of nature, something which remains an important topic in modern epistemology and philosophy of science. In the context of modern physical cosmology, some analogous ideas—e.g. the Davies-Tipler argument against cosmologies with past temporal infinity—have been proposed only relatively recently, in the last quarter of the 20th century (Tipler 1982; Ćirković 2012). This confluence of contemporary philosophy of science (in particular philosophy of cosmology and, to a smaller degree, philosophy of evolutionary theory) and the history of ancient philosophy is, to a large degree, a two-way street. While recognizing precursors of the modern anthropic reasoning helps the subject in gaining legitimacy, it is often denied by conservatives; nevertheless, the entire process could help resolve some long-standing puzzles in the history of ideas as well. Some examples in this sense include the role of uniformitarian thinking in cosmology (Balashov 1994), or the intellectual roots of the search for unifying laws (Nicolaides 2020). There are many other, potentially bold and innovative research programs to be found at the intersection of philosophy and fundamental physics and cosmology, inspired by ancient thought.

There is a great deal of intellectual inertia to be overcome on this road as well. When "a librarian of genius" in Borges's *Library of Babel* discovers the fundamental law of the Library (that it contains all variations of the finite number of typographic symbols), such an achievement is probably accompanied by various kinds of conservative resistance (Borges 1999). This Newton or Einstein of the fictional world should not have been bothered by complaints that the discovery is "metaphysical", "abstract", "ugly", not to mention "impractical". Arguably, the discovery of the fundamental law need not change anything in the daily routine of most inhabitants of Borges' universe, although it does, in a sense, explains them. Local political and administrative structures could elect to ignore it entirely.

All that does not, however, subtract from the importance of the discovery—if anything, it adds to the magnitude of the required intuitive leap and the elegance of the solution.

Few Borgesian lessons, however, could not be applied to the real world. While the architecture of *our* multiverse and the anthropic reasoning as the best tool in studying that architecture is likely to be the topic of much work in the centuries to come, a possibility that the quest will result in unexpected side benefits and resolutions should not be discounted. As concluded by Roush: "[T]he fact that a way of going about natural science mentions human beings is not sufficient reason to think that it is a subjective approach; in fact, it may *need* to mention human beings in order to be objective" (Roush 2003, 5). We may be witnessing the emergence of an improved cosmological paradigm for the third millennium: the challenge is to *not* see it necessarily in the rear-view mirror, but to be a general before the proverbial *after the battle*.

Bibliography

Balashov, Yu. 1994. "Uniformitarianism in Cosmology: Background and Philosophical Implications of the Steady-State Theory". *Studies in History and Philosophy of Science*, Vol. 25B: 933-958.

Barrow, John D. and Dabrowski, Mariusz P. 1995. "Oscillating Universes". *Monthly Notices of the Royal Astronomical Society*, Vol. 275: 850-862.

Borges, Jorge. Luis. 1999. *Collected Fictions*, translated by A. Hurley. New York: Penguin.

Brown, Geoffrey. 1984. "The Cosmological Theory of Empedocles". *Apeiron: A Journal for Ancient Philosophy and Science*, Vol. 18: 97-101.

Burnet, John. 1892. *Early Greek Philosophy*. London: Adam and Charles Black.

Caldwell, Robert R., Kamionkowski, Marc, and Weinberg, Nevin N. 2003. "Phantom Energy: Dark Energy with $w < -1$ Causes a Cosmic Doomsday". *Physical Review Letters*, Vol. 91: 071301 (4 pp).

Carr, Bernard. (Ed.). 2007. *Universe or Multiverse*. Cambridge: Cambridge University Press.

Carroll, Sean M. 2006. "Is Our Universe Natural?". *Nature*, Vol. 440: 1132-36.

Ćirković, Milan M. 2004. "The Anthropic Principle and the Duration of the Cosmological Past". *Astronomical and Astrophysical Transactions*, Vol. 23: 567-597.

Ćirković, Milan M. 2012. The Astrobiological Landscape: Philosophical Foundations of the Study of Cosmic Life. Cambridge: Cambridge University Press.

Ćirković, Milan M. 2016. "Anthropic Arguments Outside of Cosmology and String Theory". *Belgrade Philosophical Annual*, Vol. 30: 91–114.

Ćirković, Milan M. and Perović, Slobodan. 2018. "Alternative Explanations of the Cosmic Microwave Background: A Historical and an Epistemological Perspective". *Studies in History and Philosophy of Modern Physics*, Vol. 62: 1-18.

Ćirković, Milan. M. and Balbi, Amedeo. 2020. "Copernicanism and the Typicality in Time". *International Journal of Astrobiology*, Vol 19, No. 2: 101-109.

Diamond, Jared. 1992. *The Third Chimpanzee*. New York: HarperCollins.

Ellis, George F. R. and Brundrit, Geoff B. 1979. "Life in the Infinite Universe". *Quarterly Journal of the Royal Astronomical Society*, Vol. 20: 37-41.

Ellis, G. F. R., Kirchner, Ulrich, and Stoeger, William R. 2004. "Multiverses and Physical Cosmology". *Mon. Not. R. Astron. Soc.*, Vol. 347: 921-936.

Graham, Daniel. 2013. Science before Socrates: Parmenides, Anaxagoras, and the New Astronomy. USA: Oxford University Press.

Gregory, Andrew. 2007. *Ancient Greek Cosmogony*. London: Duckworth.

Guzman, Melissa, Hein, Andreas M., and Welch, Chris. 2017. "Extremely Long-Duration Storage Concepts for Space". *Acta Astronautica*, Vol. 130: 128-136.

Hoyle, Fred. 1983. *The Intelligent Universe*. London: Michael Joseph Limited.

Hoyle, Fred, Burbidge, Geoffrey R. and Narlikar, Jayant V. 1993. "A Quasi-Steady State Cosmological Model with Creation of Matter". *Astrophys. J.*, Vol. 410: 437-457.

Hoyle, Fred, Burbidge, Geoffrey R. and Narlikar, Jayant V. 1994. "Astrophysical Deductions from the Quasi-Steady-State Cosmology". *Mon. Not. R. astr. Soc., Vol.* 267: 1007-1019.

Ierodiakonou, Katerina. 2011. "Remarks on the History of an Ancient Thought Experiment". In *Thought Experiments in Methodological and Historical Contexts*, edited by K. Ierodiakonou and S. Roux, 37-49. Leiden: Brill.

Jarvis, William. 2003. *Time Capsules: A Cultural History*. North Carolina: McFarland & Company.

Kragh, Helge. 1996. *Cosmology and Controversy*. Princeton: Princeton University Press.

Kragh, Helge. 2007. Conceptions of Cosmos from Myth to the Accelerating Universe: A History of Cosmology. Oxford: Oxford University Press.

Kragh, Helge. and Longair, M. (Eds.). 2019. *The Oxford Handbook of the History of Modern Cosmology*. Oxford: Oxford University Press.

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

Lucretius 1997. *On the Nature of Things*. translated by William E. Leonard. Urbana: Project Gutenberg. Available at: http://onlinebooks.library.upenn.edu/webbin/gutbook/lookup?num=785.

Milanković, Milutin. 1956. "The Tower of Babel of Modern Technology". *Glas SAN*, Vol. 220: 137–149 (in Serbian).

Nicolaides, Demetris. 2020. In Search of a Theory of Everything: The Philosophy Behind Physics. Oxford: Oxford University Press.

North, John. 1965. The Measure of the Universe: A History of Modern Cosmology. London: Oxford University Press.

O'Brien, Denis. 1969. *Empedocles' Cosmic Cycle*. Cambridge: Cambridge University Press.

Rahman, Tahir. 2009. “How One Tiny Silicon Disc Took the Messages of Mankind to the Moon”. *Silicon*, Vol. 1: 203-206.
Rescher, Nicholas. 1985. “Extraterrestrial Science”. In *Extraterrestrials, Science and Alien Intelligence*, edited by E. Regis, Jr., 83-116. Cambridge: Cambridge University Press.
Rose, Christopher and Wright, Greg. 2004. “Inscribed Matter as an Energy-Efficient Means of Communications with an Extraterrestrial Civilization”. *Nature*, Vol. 431: 47-49.
Rosenmeyer, Thomas G. 1989. *Senecan Drama and Stoic Cosmology*. Berkeley: University of California Press, Berkeley.
Roush, Sherrilyn. 2003. “Copernicus, Kant, and the Anthropic Cosmological Principles”. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, Vol. 34: 5-35.
Tegmark, Max. 1998. “Is ‘The Theory of Everything’ Merely the Ultimate Ensemble Theory?”. *Annals of Physics* Vol. 270: 1-51.
Tegmark, Max. 2003. “Parallel Universes”. *Scientific American*, Vol. 288: 40-51.
Tegmark, Max. 2008. “The Mathematical Universe”. *Foundations of Physics*, Vol. 38: 101-150.
Thucydides. 1843. *History of the Peloponnesian War*. Translated by Thomas Hobbes. London: Bohn. Available at: http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0247%3Abook%3D1.
Tipler, Frank J. 1982. “Anthropic-Principle Arguments Against Steady-State Cosmological Theories”. *The Observatory*, Vol. 102: 36-39.
Vilenkin, Alexander. 2006. *Many Worlds in One*. New York: Hill and Wang.
Weinberg, Steven. 1972. Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity. Hoboken: Wiley-VCH.
Weinberg, Steven. 2008. *Cosmology*. Oxford: Oxford University Press.

CHAPTER TEN

EXPLANATORY ROLE OF THE PRINCIPLE OF CAUSAL RESEMBLANCE IN EARLY IONIAN PHYSICS: ARISTOTLE'S ACCOUNT

DRAGO ĐURIĆ

I. Introduction

The *principle of causal resemblance* (Makin 1991) coincides with or overlaps with the *transmission theory of causation* (Lloyd 1976; Mourelatos 1984), the *principle of causal synonymy* (Aristotle 1933;[1] Barnes 1982; Hankinson 1998), and the *principle of like causes like* (Marmodoro 2017). Generally speaking, the principle and its related theories tell us that the cause must in some sense resemble its effect. The various formulations result from the fact that different thinkers have focused on different aspects of the causal relations. Descartes points to some of these aspects when he says: "But now it is manifest by the natural light that there must be at least as much reality in the total and efficient cause as in its effect. For, I ask, from where could the effect derive its reality, if not from the cause? And how could the cause give it reality, if it did not also possess it?" (Descartes 2008, 29).[2] In all likelihood, Descartes is thinking of all four types of Aristotle's causes (material, formal, efficient, and teleological). It sometimes seems that Aristotle believes this principle applies to all of them. But even though he provides examples of the application of the principle to the other

[1] For other places where Aristotle considers this principle, see Mourelatos 1984.

[2] As a purified scholastic formulation of the principle, Gilson cites the formulation of Eustachio a Sancto Paulo, which reads "nihil est potest in effectu quod non praeextiterit in causa" (Eustachio a Sancto Paulo 1609, 3.428.1-2). For other, similar formulations of this principle, see Gilson 1979, 44-45.

three causes, its application to material causes seems the most convincing.[3] This is particularly relevant in his consideration of the material cause of early Ionian physicists (Thales, Anaximander, Anaximenes, Heraclitus).

In what follows, I will defend the thesis that without the assumption of at least implicit respect for what this principle requires, it is difficult to justify or even understand the motivation for the materialistic theories of these Ionian physicists. My consideration is based on Aristotle's conceptual framework and his interpretive model of material monism (Thales, Anaximenes, Heraclitus) and material pluralism (Anaximander). I will show that the ultimate goal of the physicists was not to determine the material content of things of the phenomenal world, but to explain the changes in them. At the end, I will defend the thesis that, for the early Ionian materialists, change in the phenomenal world is explained by changes in the fundamental substance; i.e. that the changes in properties of things of the phenomenal world are only epiphenomena of the changes in relations between the units of the fundamental substance.

II. Aristotle's Conceptual Network[4]

I will now say something about the three basic terms of Aristotle's philosophy that are important for his interpretation of earlier philosophies: αἰτίων, ἀρχὴ, and στοιχεῖον. In *Posterior Analytics*, he says: "We suppose ourselves to possess unqualified (*ἁπλῶς*) scientific knowledge of a thing, as opposed to knowing it in the accidental way in which the sophist knows, when we think that we know the cause (αἰτίαν) on which the fact depends, as the cause of that fact" (Aristotle 1960, 71b 9-11).[5] But philosophy or the highest wisdom is not just any knowledge—it is the highest knowledge. To gain such knowledge, it is not sufficient to know the closest or, in fact, any cause in the causal chain; it is necessary to reach the first or ultimate cause. Aristotle expresses this ultimate cause by the word ἀρχὴ. In the *Delta* book of his *Metaphysics*, he defines the common (κοινὸν) meaning of the word ἀρχὴ as "the first thing (τὸ πρῶτον) from which something either exists or

[3] In his objection (Fifth Objections and Replies) to Descartes, Gassendi says: "The causal principle really applies to the material cause, not the efficient cause, which can be of a very different nature to the effect and cannot be said to give its own reality to the effect" (Descartes 2008, 172).

[4] I will leave aside the question of the reliability of the image Aristotle suggests to us about early Ionian physics. For this, see Cherniss 1935; Guthrie 1957; Kahn 1960; Stevenson 1974; Collobert 2002; and Gomes 2016.

[5] Aristotle, *Apo*, 71b 9-11. /= Aristotle. *Analitica posteriora.* In Aristotle. *Posterior Analytics. Topica.*

comes into being or becomes known (ἢ γιγνώσκεται)" (Aristotle 1933, 1013a 18-19). In the *Alpha* book of *Metaphysics*, Aristotle says philosophy "must obtain knowledge of the primary causes, because it is when we think that we understand its primary cause (πρώτην αἰτίαν) that we claim to know each particular thing" (Aristotle 1933, 983a 24-26).

He then lists the four types of ultimate causes mentioned previously: material, efficient, formal, and final. The ultimate material cause or material ἀρχὴ, he says, is "that of matter or substrate (τὴν ὕλην καὶ τὸ ὑποκείμενον)" (Aristotle 1933, 983a 30-31). He often calls this type of cause στοιχεῖον. In the *Delta* book of *Metaphysics*, Aristotle defines στοιχεῖον[6] as "the primary immanent thing, formally indivisible into another form, of which something is composed", things which are "not farther divisible", and "the parts into which bodies are ultimately divisible". For present consideration, it is important to add that, according to Aristotle, the word στοιχεῖον also means "any small unity that is usuful for different purposes", that is to say, what is "small or simple or indivisible" (Aristotle 1933, 1014a 26-1014b 6). While Aristotle uses the terms ἀρχὴ, στοιχεῖον, αἰτίων and ὕλη for ultimate material and material cause, I will continue, for economy of explication, to use the term fundamental substance.

III. Traditional Speculations on the Motive for Monism with Respect to the Ultimate Cause

In this part of my paper, I show that the ultimate motive for material monism or pluralism is not to explain the ultimate matter from which the various things of a phenomenal world are made, and which then consist of it, but to explain the changes in this world. I will defend the thesis that the first physicists came to this matter, as well as to its properties, in an effort to explain these changes. To be eligible for ἀρχὴ, the matter must be such that it makes changes in the phenomenal world possible, or that, at least, it does not make them impossible.

Before I continue, one more problem needs to be considered. Presocratic theories are usually divided into monistic and pluralistic theories. However, it should be kept in mind, and this is often not the case, that there are two criteria for this division. On the one hand, the number of kinds of ultimate causes is a concern. We must distinguish between monistic theories, in which all the changes in the world are explained with one kind of cause, as do the first Ionian materialists; and pluralistic theories, in which these early

[6] For details on the meaning of the word στοιχεῖον in ancient Greek, see Diels 1899.

physicists, in the explanation of the changes, in addition to the material cause, introduce at least one other kind of cause, for example, the efficient cause. On the other hand, Aristotle's interpretation implies a division in terms of the number of causes within the one kind of ultimate cause. According to this interpretation, all early Ionian physicists are causal monists, but some of them are material monists (Thales, Anaximenes, and Heraclitus), and others are material pluralists (e.g. Anaximander).

Most authors who suggest a possible motivation for monism have material monism and not material pluralism in mind. Let us look at some of these attempts. Burnet, for example, says these philosophers "could find no satisfaction in the view of the world as a perpetual contest between opposites", they searched for "something which persisted through all change, and ceased to exist in one form only to reappear in another" (Burnet 1920, 6). Burnet, unfortunately, does not disclose any scientific or philosophical reasons for the search for something that persists through all changes, or under what conditions this something can persist through all changes. Although the second claim, namely that this something "ceased to exist in one form *only* to reappear in another", suggests a kind of teleological explanation, it is difficult to find a testimony to support such a theory.

Guthrie says Milesians were "the first to attempt on a rational basis that simplification of reality which has been the quest of the human mind in all ages", and their monism was the result of a kind of "impulse to simplification" (Guthrie 1962, 57). In short, he seems to think monism is a consequence of the "simplification of reality", and the "simplification of reality" is the result of satisfying the "impulse to simplification". At this point, Guthrie relies on Stebbing, who says that "there seems to be a deep-rooted tendency in the human mind to seek /…/ something that persists through change" (Stebbing 1933, 404). According to this interpretation, monism is the result of a psychological tendency ("a deep-rooted tendency in the human mind"). It has neither philosophical nor scientific origins.

Barnes says monism is "the simplest hypothesis that will account for the constitution of the world: unity is simpler than plurality." He then concludes that "science always strives for economy and simplicity in explanation", adding that by postulating water as the one and only matter from which everything is composed, Thales became an "embryonic scientist", for he saw, writes Barnes, that it was "eminently simple and because of its simplicity he adopted it as a hypothesis" (Barnes 1982, 11). And here, the ultimate motive for monism is the fact that "science always strives for economy and simplicity in explanation". In a similar way, Feinberg says Thales' monism answers the question "of the structure and composition of bulk matter, by which is meant the objects and substances we find around

us". He concludes that "already in the asking of this question there is implicit the assumption of a simplicity underlying the complexities of bulk matter" (Feinberg 1966, 6).

Some of these (and many similar) speculations about possible motives for a monistic hypothesis suggest that monists were looking for simplicity. It is strange that in the above considerations, in addition to some kind of personal psychological motive for simplification ("impulse to simplification", "a deep-rooted tendency in the human mind" for the "simplification of reality"), there is also some kind of scientific psychological motivation ("science always strives[7] for economy and simplicity"). Some of the interpreters emphasize the importance of the opposition of the simplicity of the underlying matter to the complexities of the phenomenal world. As can be seen, the motive for simplicity in this case is that it explains "the structure and composition of bulk matter" of the phenomenal world, but nothing is said about why we would search for "the structure and composition of bulk matter" in the first place. It is also said to be a motive of a monist to find something simple that "persists through all change" (Burnet, Barnes), but no clear reasons are given for this.

If the motive of the earliest physicists was only to explain the "structure and composition of bulk matter", i.e. "the objects and substances we find around us", then it would have been much easier to use material pluralism: different things in our phenomenal world have different properties thanks to the various substances from which they are made. Yet the answer to the question of the motives for material monism or material pluralism requires more than answering the question of why these thinkers were searching for one or more fundamental substances; rather, we must consider why they would search for simple fundamental substances at all.

In his reflections on the teachings of the early Ionian physicists, Aristotle frequently seems to resort to phenomenological interpretation of the fundamental substance. But he also notes that those who choose one material ἀρχὴ, belonging to the phenomenological world, have a problem with opposites or opposing properties. These are, as Aristotle says in *Physics*, "already intertwined with pairs of contraries" (μετ' ἐναντιοτήτων συμπεπλεγμένα ἐστίν) (Aristotle 1957, 189a 2). It seems that he had in mind the following: if for a material ἀρχὴ, which is boundless and whose properties are immutable, we choose fire, then all things of the world would be hot. If we choose water, then all things would be wet, etc.[8]

[7] Here we are dealing with the personification of science itself ("science ... strives").
[8] Aristotle says a little later on that these substances are, by their properties, "opposite to each other, since, for example, the air is cold, the water is moist, while

Obviously this is not the case for the things of our world. However, if we take four classical elements, then the most suitable for ἀρχὴ, thinks Aristotle, is air. Why? Because, air "presents sensible differences (διαφορὰς αἰσθητάς) in a less degree than the others" (Aristotle 1957, 204b 27-28). With this, he suggests the material ἀρχὴ should have minimal properties. In fact, it would be best if it did not have any properties at all. Such an ἀρχὴ could somehow persist through all the changes, without leading to the "interweaving" of opposites or opposite properties.[9] However, if material ἀρχὴ is defined as something that cannot have the properties of things of the phenomenal world, why do Thales, Anaximenes, and Heraclitus give their ἀρχὴ the names of things of our phenomenal world: water, air, fire? I will return to this later. For now, suffice it to say that either way, the above consideration leads us to the conclusion that, regardless of changes in properties in the phenomenal world, at a fundamental level, a material cause must resemble a material effect.

IV. Changes in the Phenomenal World Cannot be Explained Without Respecting the Principle of Causal Resemblance

Although some interpreters have recognized the principle of causal resemblance in the testimonies of Alcmeon's thought (Barnes, Hankinson), I am inclined to believe that a persuasive reason for adhering to this principle cannot be traced to Anaxagoras' consideration of the following problem: "For how can hair come from what is no hair or flesh from what is not flesh" (Πῶς γὰρ ἄν', φησίν, 'ἐκ μὴ τριχὸς γένοιτο θρὶξ καὶ σὰρξ ἐκ

the fire is hot" (πρὸς ἄλληλα ἐναντίωσιν, οἷον ὁ μὲν ἀὴρ ψυχρός, τὸ δ' ὕδωρ ὑγρόν, τὸ δὲ πῦρ θερμόν) (Aristotle 1957, 204b 27-28).

[9] This is one of the main points of Plato's consideration of the early physicists presented in Plato, *Tim*., 49a-51b. The most commonly encountered opposites (ἐναντία) are cold/ hot and wet/ dry. Obviously, this is something we usually call properties today. Earlier thinkers called them opposites because nothing can have both properties at the same time. No one thing can be at the same time, for example, hot and cold or wet and dry. Aristotle, in *GC*, points to the problem in the following way: "Contraries, however, refuse to be coupled; for it is impossible for the same thing to be hot and cold, or moist and dry" (Aristotle, *GC* 330a 31-33 /= Aristotle. *De generatione et corruptione*. In Aristotle. *On sophistical refutations. On coming-to-be and passing away. On the cosmos*.

μὴ σαρκός') ()?[10] It is a fact that hair grows, for example, from the flesh of the head, although the properties of hair and the properties of flesh are radically different. Yet this fact seems inexplicable only if we endorse the following two assumptions: 1) the properties of the effect must resemble the properties of the cause, and 2) these causes and effects are the things of our phenomenal world, with properties available to our senses. It seems Anaxagoras would not ask the above questions, nor would he wonder about such or similar phenomena, if he did not think these changes are contrary to something like the principle of causal resemblance.

It is quite possible that Anaxagoras was not the first to formulate this principle. It is even less likely that this principle was a special topic in his reasoning. But he could reason in accordance with it without special emphasis on it or creating terms for it.[11] He could see that the things of our world that become from other things resemble them and inherit their properties in some way. A statue made of stone looks like a stone, and gold jewellery looks like gold. But he could also see, for example, that a calf enlarges its body mass (flesh) by eating the grass, and flesh becomes from grass, although grass, with respect to properties, does not resemble flesh at all.

However, even in those processes in which one thing becomes from some other thing of similar properties, he could perceive at least a slight difference in the properties between them. He could have noted the following: although the thing *y*, which we ourselves make, consists of material taken from some other thing *x* (or several other things), it does not inherit all their properties. These two things must be at least minimally different. Without a minimal difference, the change cannot be registered. After all, the change implies a difference in properties between the condition of a thing before and after.

Anaxagoras, in all likelihood, is trying to solve the following riddle: how does a change that appears to be impossible occur? I suggest that, according to Aristotle, this is also the case for the early Ionian physicists. This riddle can be seen in Pseudo-Aristotle's testimony, in which the Eleatic style of the argument for the impossibility of becoming is explicated. This argument is attributed to Xenophanes, although the authenticity of the testimony is challenged (see Mansfeld 1988 and Finkelberg 1990). It goes as follows:

---

[10] Anonym. *Scholia in S. Gregorii Orationes*. In *Patrologia Graeca* 36, (Paris: Migne, 1853), 911, DK B10 /= Hermann Diels and Walther Kranz. *Die Fragmente der Vorsokratiker*. (Zurich/Hildesheim: Weidmann, 1985)/.

[11] For more about this, see Stannard 1965. People act in accordance with the principle of non-contradiction, even if they do not know such a principle exists.

> He says that, if enything exists, it cannot have become /…/ For that which has come into existence must have become either from like[12] or from unlike. But neither of these is possibe (᾽Αδύνατόν φησιν εἶναι, εἴ τι ἔστι, γενέσθαι … ἀνάγκη γὰρ ἤτοι ἐξ ὁμοίου ἢ ἐξ ἀνομοίου γενέσθαι τὸ γενόμενον· δυνατὸν δὲ οὐδέτερον). (Aristotle [Pseudo] 1963, DK A28)

If something becomes from the like, then there is no becoming. If *x* is a thing that should be the material cause of thing *y*, and if *y* in respect to its properties is absolutely matched with *x*, then *y* cannot have become, for it already exists. However, *y* cannot become from a thing *z*, which does not have the properties of *y*, since, according to Descartes's definition of the principle, the cause must have everything necessary for the effect.

Anaxagoras solves this problem by introducing the principle by which a *portion of everything is contained in everything*.[13] If we accept this principle, it is not difficult to conclude that everything that is needed for the growth of hair is contained in the skin. In this way, Anaxagoras saves *the principle of causal resemblance* by appealing to the principle according to which a *portion of everything is contained in everything*. However, not only can we not, with our senses, see all things in all things, but we cannot see them in any of the things of our phenomenal world. But how does he come to the conclusion that all things are present in all things? The answer to this question, in all likelihood, cannot be found in the testimonies of his philosophy. We can, for example, imagine that his thinking resonated with some kind of rudimentary form of argumentation, which could roughly look like this:

1. On the basis of our sensory experience, we know there are changes in the things in the phenomenal world; from one thing (or from its parts or from several things or parts of several things), another thing (or more) becomes, which differs from the original in terms of properties.
2. These changes would not be possible if both things were not made up of the same or similar fundamental substance or of several substances, although its/their existence is not available to our senses.
3. So there is some such fundamental substance, although its existence is not available to our senses.

[12] In Liddell, Scott, Jones et al. (1996), ὅμοιος is translated, inter alia, as "like", "resembling" and "same".

[13] Testimonies that Anaxagoras advocated this principle are found in almost all DK B fragments. It is interesting that the fragments are taken from Simlicius' commentary on Aristotle's Physics. See, for example Simplicius 1882, (DK B1), 164.20 (DK B3), 24.26, (DK B4b), 164.27 (DK B6), 175.12 (DK B8) etc.

Anaxagoras's solution assumes there is a difference between the phenomenal and the fundamental or archaeological level or aspect of physical reality. At the phenomenal level, the skin and hair differ with respect to their proprieties, but they do not differ at the fundamental level. Therefore, unlike the Pseudo-Xenophanes's argument given above, things can become from both like and unlike things, but only at the phenomenal level, while at the fundamental level, they must always, in some sense, be like. Anaxagoras arrives at his solution on the fundamental material content of the things in an effort to explain the changes in the phenomenal world.

Aristotle's interpretation of early Ionian materialists also suggests that their ultimate goal is to explain changes in the phenomenal world. What satisfies the conditions of being a fundamental substance depends crucially on the theory of change endorsed by each of these thinkers. Aristotle distinguishes two types of change theories:

> According to what physicists say, there are two ways [of explanation of changes of things]. The first set make the underlying body one of either one of the three [water, air, fire] or something else which is denser than fire and rarer than air, then generate everything else from this, and obtain multiplicity by condensation and rarefaction (ὡς δ' οἱ φυσικοὶ λέγουσι, δύο τρόποι εἰσίν. οἱ μὲν γὰρ ἓν ποιήσαντες τὸ [ὂν] σῶμα τὸ ὑποκείμενον, ἢ τῶν τριῶν τι ἢ ἄλλο ὅ ἐστι πυρὸς μὲν πυκνότερον ἀέρος δὲ λεπτότερον, τἆλλα γεννῶσι πυκνότητι καὶ μανότητι πολλὰ ποιοῦντες). (Aristotle 1957, 187a 13-17).

Therefore, if the material ἀρχή, τὸ ὑποκείμενον, or the fundamental substance is one body (Thales' water, Anaximenes' air, Heraclitus' fire, or some body denser than fire and rarer than air), the generation and change occur through condensation and rarefaction. However, some physicists, among whom Aristotle includes Anaximander, do not agree. Aristotle says:

> The second set [of physicists] assert that the opposites are contained in the one and emerge from it by separation, for example, Anaximander and also all those who assert that 'what is' is one and many, like Empedocles and Anaxagoras; for they too produce other things from a mixture by separation (οἱ δ' ἐκ τοῦ ἑνὸς ἐνούσας τὰς ἐναντιότητας ἐκκρίνεσθαι, ὥσπερ Ἀναξίμανδρός φησι, καὶ ὅσοι δ' ἓν καὶ πολλά φασιν εἶναι, ὥσπερ Ἐμπεδοκλῆς καὶ Ἀναξαγόρας· ἐκ τοῦ μίγματος γὰρ καὶ οὗτοι ἐκκρίνουσι τἆλλα). (Aristotle 1957, 187a 20-23).

Therefore, according to his theory, the opposites or properties are separated from the one that should be their fundamental substance, or material ἀρχή. This one is not, as in the case of the first physicists, a single

body (ἓν ... σῶμα), a homogeneous mass, but a mixture[14] from which all things of phenomenal world become. While in the first case, all things of the phenomenal world become by condensation and rarefaction of the fundamental homogeneous substance; here other things become by separation from that one, which is a mixture and therefore heterogeneous. More light can be shed on this division by a very similar division mentioned by Aristotle at the beginning of his book *On Coming-to-be and Passing-away*:

> Those, then, who construct all things out of a one [ultimate material cause] must maintain that coming-to-be and passing-away are alteration. For they hold that the underlying something always remains same and one; and we call such a change alteration. Those who hold that ultimate kinds of things are more than one, maintain that alteration is distinct from coming-to-be: for coming-to-be and passing-away correspond to the joining and the separation (Τοῖς μὲν οὖν ἐξ ἑνὸς πάντα κατασκευάζουσιν ἀναγκαῖον λέγειν τὴν γένεσιν καὶ τὴν φθορὰν ἀλλοίωσιν· ἀεὶ γὰρ μένειν τὸ ὑποκείμενον ταὐτὸ καὶ ἕν· τὸ δὲ τοιοῦτον ἀλλοιοῦσθαί φαμεν· τοῖς δὲ τὰ γένη πλείω ποιοῦσι διαφέρειν τὴν ἀλλοίωσιν τῆς γενέσεως· συνιόντων γὰρ καὶ διαλυομένων ἡ γένεσις συμβαίνει καὶ ἡ φθορά). (Aristotle 1955, 314b 2-7)

The above classifications can be summarized in the following way:

1. Material monists: a) in respect to its proprieties, ἀρχή is a homogeneous fundamental substance; b) change in every thing of the phenomenal world depends on condensation and rarefaction of the fundamental substance, and this condensation and rarefaction emerges at the phenomenal world as c) an alteration.
2. Material pluralists: a) with respect to its properties, ἀρχή is a mixture of several different (in some cases, an infinite number) ultimate material causes; b) every thing of the phenomenal world becomes and disappears by separating from that mixture, and this separation (coming together) of different substances that makes up a fundamental mixture c) results in a change in the things of the phenomenal world.

Thus, the changes of the properties of things of the phenomenal world that take place through condensation or rarefaction of the fundamental substance are consistent with material *monism* (Thales, Anaximenes,

[14] Aristotle also claims in *Metaphysics* that Anaximander's fundamental substance is a "mixture (τὸ μῖγμα)" (Aristotle 1933, 1069b 22). See Aristotle's consideration of the problem of mixture in: Aristotle 1955, 1.10. The way in which it might be brought about in relation to his understanding of Anaximander's material pluralism would require special exploration.

Heraclitus), while the changes that take place through separation (and coming together) are consistent with material *pluralism* (Anaximander).

Both change theories imply the existence of something like elementary units of a fundamental substance and something commonly called empty space.[15] Without this assumption, it is not easy to imagine processes of condensation or rarefaction or of separation and coming together, which, according to these theories, occur in that substance. But if changes in the phenomenal world are only the consequences of these processes, it is not easy to see why the fundamental substance would be, for example, air and not water or fire. If we have material monism in mind, then water, air, and fire differ only in their degree of condensation or rarefaction. But the question arises: what actually becomes denser or rarer? It seems there may be no water, no air, no fire, but simply some units of the fundamental substance from which water and air and fire are made.

If we accept the assumption that the ἀρχή or στοιχεῖον, in terms of its properties, is immutable, then none of the things of the phenomenal world can be an ἀρχή or στοιχεῖον, since none can "persist through all changes". If rarefaction of water produces air, then, according to Graham's interpretative model of the generating substance theory, "it [water] ceases to exist and is replaced by a successor substance [air]" (Graham 2006, 70). Water and air, as we know, have different properties. For this reason, for example, air as a substance of the phenomenal world cannot remain gaseous if it is transformed into water or stone. In that case, the material cause would not resemble the material effect, and, thus, the principle of causal resemblance would not be respected. If we want to save the thesis that, for example, air is a fundamental substance, then we might speculate that early Ionian physicists had in mind some fundamental air that would differ from phenomenal air. But another question arises: why call it air at all?

---

[15] Based on some verses in Parmenides's poem where he seems to criticize the theory of the early Ionian physicists, it could be concluded that they had some idea of empty space. Without this idea, it is not easy to imagine the processes of condensation and rarefication of the fundamental substance. Namely, Parmenides argues: "For neither not being [what is not] exists, which would stop it [being] from reaching the same [being], nor is there any way in which being would be more here and less there (Οὔτε γὰρ οὐκ ἐὸν ἔστι, τό κεν παύοι μιν ἱκνεῖσθαι εἰς ὁμόν, οὔτ' ἐὸν ἔστιν ὅπως εἴη κεν ἐόντος τῇ μᾶλλον τῇ δ' ἧσσον)"(*DK* B8.46-49a).
Based on the above statements, Parmenides concludes that being exists "everywhere equal to itself". To "be more here and less there" can mean to be denser here and rarer there, and this difference in the degree of condensation/rarefication of the fundamental substance causes the beings of the phenomenal world to be unequal. Apparently, because of these logical problems (to say that "not being [empty space] exists"), the non-being or empty space was positively expressed by atomists.

For now, it seems the only way to preserve the principle of causal resemblance is to imagine a fundamental substance consisting of elemental units whose properties are unchangeable. We might imagine that early Ionian physicists had in mind the cosmogony according to which, before the becoming of the world, the entire fundamental substance was in a certain degree of condensation or rarefaction, which at the phenomenal level would look like water or air or fire. Thales could go from what seemed to him the most widespread substance in the world and conclude that before the world became, there was only water, or a distinctive degree of condensation of the fundamental substance. Anaximenes could have thought of air in a similar way. Other diverse things of our world could have become in such a way that only a small part of this original substance (by condensation or rarefaction of an elementary unit of a fundamental substance) was transformed into them.

Heraclitus's motive for the choice of fire as a fundamental substance could be different. One of the most important testimonies of his philosophy tells us that he saw the world as eternal.[16] This should exclude the possibility of any cosmogony. He could have in mind the thesis that the fundamental substance should be something that is the rarest, and that thing, according to Anaximenes, is fire. Aristotle's observation from the *Alpha* book of *Metaphysics* goes in the same direction:

> [T]he most elementary body is that from which first other bodies are produced by combination; and this will be that body which is rarest and composed of the finest particles. Hence all who posit fire as a first principle will be in closest agreement with this theory.[17] However, even among the other thinkers everyone agrees that the primary corporeal element is of this kind. At any rate none of the monists thought earth likely to be an element—obviously on account of the size of its particles—but each of the other three has had an advocate; for some name fire as the primary element, others water, and others air (δόξειε στοιχειωδέστατον εἶναι πάντων ἐξ οὗ γίγνονται συγκρίσει πρώτου τοιοῦτον δὲ τὸ μικρομερέστατον καὶ λεπτότατον ἂν εἴη τῶν σωμάτων (διόπερ ὅσοι πῦρ ἀρχὴν τιθέασι, μάλιστα ὁμολογουμένως ἂν τῷ λόγῳ τούτῳ λέγοιεν: τοιοῦτον δὲ καὶ τῶν ἄλλων ἕκαστος ὁμολογεῖ τὸ

[16] According to the testimony, Heraclitus claims the following: "This world-order, the same of all, no god nor man did create, but it ever was and is and will be everliving fire" (Clemens Alexandrinus 1905-1936, 5.103.6; DK B30.

[17] In *Physics*, Aristotle says the best candidate for the material ἀρχή is air (Aristotle, 1957, 189a 5), while here he says it is fire. Note that Aristotle favors two different candidates for ἀρχή by starting from different criteria. In the first case, the material ἀρχή must be something that has the least properties, and in the second, that which is the rarest.

στοιχεῖον εἶναι τὸ τῶν σωμάτων: οὐθεὶς γοῦν ἠξίωσε τῶν ἓν λεγόντων γῆν εἶναι στοιχεῖον, δηλονότι διὰ τὴν μεγαλομέρειαν, τῶν δὲ τριῶν ἕκαστον στοιχείων εἴληφέ τινα κριτήν, οἱ μὲν γὰρ πῦρ οἱ δ᾽ ὕδωρ οἱ δ᾽ ἀέρα τοῦτ᾽ εἶναί φασιν). (Aristotle 1933, 988b 35-989a 8)

Aristotle tells us that these physicists were searching for some substance that is the most elemental (στοιχειωδέστατον), either something made up of the smallest parts (τὸ μικρομερέστατον) or elementary particles that are finest or smallest with respect to quantity (λεπτότατον). He believes no one has chosen earth as a fundamental substance because of the greatness of its parts (μεγαλομέρειαν).[18] Since all other things of the phenomenal world owe their properties to the degree of density of the particles or elementary units of the fundamental substance, Aristotle seems to suggest densification is not a proportional reduction of the distance between all elementary units of some thing of which the phenomenal world consists; rather, it could represent the formation of smaller or larger conglomerates of these elementary units. Therefore, we could imagine a theory according to which fire would consist of separate units of a fundamental substance, denser things from a conglomerate of two such units, even denser things from three etc. A similar theory could be conceived from material pluralism. This theoretical model would be very similar to modern chemistry where molecular changes occur as a result of different combinations of unchangeable atoms. In this way, keeping in mind the elemental units of the fundamental substance (*στοιχεῖα)*, we could preserve the principle of causal resemblance. From the way these early physicists imagined unchangeable ingredients of the fundamental substance, atomistics could conclude they are also unchangeable in terms of quantity. "The most elemental (στοιχειωδέστατον) substance", or the substance which consists of "the smallest parts (τὸ μικρομερέστατον)", is not divisible into smaller parts.

V. Some Other Implications of Aristotle's Consideration of Ionian Physicists' Theories of Change

Now, a common characteristic of both theories of change is that they take place as a mutual approximation and removal of the unchangeable parts (*στοιχεῖα*) of the fundamental substance. Material monism suggests an approximation and removal of the parts of the fundamental substance with

[18] It is obvious that the material ἀρχή or στοιχεῖον must be flexible; that is, it must be fit for changing things in the phenomenal world. This can be better understood by looking at what the word στοιχεῖον means to Aristotle. See Aristotle 1933, 5.3 114a 26 – 114b 15.

the same properties, and material pluralism suggests the parts of the fundamental substance have different properties. If a thing *x* of phenomenal world, with a set of properties *A*, changes to a thing *y* with a different set of properties *B*, then the principle of causal resemblance is violated. However, if *x* and *y* at the fundamental level consist of one and the same substance (water, air, fire), then this principle is not violated, and the change can be explained. According to material pluralists, *στοιχεῖα* of the fundamental substance participate in the properties of the phenomenal world by bringing in and reliving their various and immutable properties. On this basis, it could be concluded that the changes in the properties of the phenomenal world are, in fact, only epiphenomena of the changes in the relations between the elementary particles of the fundamental substance.

Bearing this in mind, we could conclude that material pluralism is more acceptable than material monism, because in the former theory, the properties of the things of the phenomenal world are inherited from the properties of the elementary units of the fundamental substance. This is not the case for the latter. In material monism, the elementary units of the fundamental substance with the same proprieties, cause, through condensation and rarefaction, changes in the proprieties of the things of the phenomenal world by means of, so to speak, "a miracle" of alteration.

The radical implication of Aristotle's interpretation of the teachings of early Ionian thinkers is not that the material cause must resemble the material effect, nor that it is, in a sense, greater or equal to it, but rather that it is numerically identical with it. Namely, if a thing consists of exactly the ingredients of which it is made (and *vice versa*), then the material cause and the material effect must represent one and the same substance. If thing *B* of the phenomenal world is made up of only a certain number of fundamental ingredients, which previously consisted of thing *A*, then the rest of the ingredients of thing *A* do not have a causal role and, therefore, are not the material cause of thing B. If we put things thusly, then, at first glance, it seems that when it comes to material causation, it is not possible to analyse the changes on the basis of the causal relations: material cause—material effect. However, it is possible that the same units of the fundamental substance can be considered in one case as the elementary ingredients of a thing of the phenomenal world, for example, thing *A*; and in the other case as the building blocks of other things, say things *B*. If, for example, air becomes from water, then the fundamental substance changes from a more condensed state (in which the fundamental ingredients play the role of cause) to a less condensed state (in which they play the role of effect). It seems that only in this way can both the principle of causal resemblance (the properties of the fundamental ingredients remain unchanged) and the

difference (the relations between them change) between material cause and material effect be preserved.

VI. Conclusion

Although we have no evidence that early Presocratic thinkers formulated the principle of causality, or that they formulated the principle *ex nihilo nihil fit*, it is thought that reasoning with these principles was a common assumption.[19] Aristotle suggests this.[20] Something similar could be argued for the principle of causal resemblance. Moreover, we might say this principle is a kind of theoretical specification of the above principles. Not only does this principle tell us that any becoming from nothing is impossible; i.e. everything that becomes must have the cause of its becoming; it also tells us, given Aristotle's understanding of, so to speak, the archaeology of these early physicists, that everything can become from everything. It was only later that Anaxagoras, if we rely on the testimonies, noticed that, in nature, not everything actually becomes from everything. Namely, to explain the fact that, for example, hair always grows from the head and never the head from the hair, a cause independent of matter must be introduced—the mind (νοῦς), which brings order into the events in the world.[21]

[19] See more about this in Mourelatos 1981. See also criticism in Gallop 1981.
[20] See, for example: Aristotle1957, 187a 27–29 and 34–35, 191a 24–31; 1933 1062b 24–25, 1032b 30–31, 983b 6–21, 984a 27–984b l.
[21] Simplicius 1882, 164.24; 156.13, DK B12.

Bibliography

Anonym. 1853. Scholia in S. Gregorii Orationes. In Patrologia Graeca, Vol. 36, Paris: Migne.

Aristotle [Pseudo]. 1963. *De Melisso*. *Xenophane*. *Gorgia*. In *Minor Works* (Aristotle), Cambridge, Mass./London: Harvard University Press/William Heinemann Ltd.

Aristotle. 1933. *Metaphysics* (Vol. I-II). Cambridge, MA: Harvard University Press.

Aristotle. 1955. De generatione et corruptione. In On Sophistical Refutations. On Coming-to-be and Passing Away. On the Cosmos. Cambridge, MA: Harvard University Press.

Aristotle. 1957. *Physics* (Vol. I-II.) Cambridge, MA: Harvard University Press.

Aristotle. 1960. *Analitica Posteriora*. In *Posterior Analytics. Topica.* Cambridge, MA: Harvard University Press.

Barnes, Jonathan. 1982. *The Presocratic Philosophers*. London: Routledge.

Barnes, Jonathan. 1984. *Complete Works of Aristotle* (2 Vols.). The Revised Oxford Translation. Princeton: Princeton University Press.

Burnet, John. 1920. *Early Greek Philosophy* (3rd ed.) London: A & C Black.

Cherniss, F. Harold. 1935. *Aristotle's Criticism of Presocratic Philosophy*. Baltimore: The John Hopkins Press.

Clemens Alexandrinus. 1905-1936. *Stromata* (4 vols. ed.) Edited by O. Stählin. Leipzig: Henrichs.

Collobert, Catherine. 2002. "Aristotle's Review of the Presocratics: Is Aristotle Finally a Historian of Philosophy?". *Journal of the History of Philosophy*, Vol. 40, No 3: 281-295.

Descartes, René. 2008. *Meditations on First Philosophy with Selections from the Objections and Replies*, translated with an introduction and notes by Michael Moriatry. Oxford: Oxford University Press.

Diels, Hermann. 1899. *Elementum*, Leipzig: Teubner.

Diels, Hermann and Kranz, Walther. 1985. *Die Fragmente der Vorsokratiker*. Zurich/Hildesheim: Weidmann.

Eustachio a Sancto Paulo. 1609. *Summa philosophiae qudripartita*. Paris: Carolus Chastellain.

Feinberg, Gerald. 1966. "Physics and the Thales Problem". *The Journal of Philosophy*, Vol. 63, No. 1: 5-17.

Finkelberg, Aryeh. 1990. "Studies in Xenophanes". *Harvard Studies in Classical Philology*, Vol. 39: 103-167.

Gallop, David. 1981. "Ex Nihilo Nihil, In Nihilum Nil: A Reply to Mourelatos". *The Journal of Philosophy*, Vol. 78, No. 11: 666-667.

Gilson, Étienne. 1979. *Index Scholastico-Cartesien*. Paris: J. Vrin.

Graham, Daniel. 2006. *Explaining the Cosmos*: *The Ionian Tradition of Scientific Philosophy*. Princeton and Oxford: Princeton Univrsity Press.

Gustavo, L. Gomes. 2016. "The Philosopher, his Predecessors, the Commentator and his Critics: on the Criticism of Herold Cherniss's Critique of Aristotle as a Source for Early Greek Philosophy". *Anais de Filosofia Classica*, Vol. 10, No. 19:76-93.

Guthrie, K. C. William. 1957. "Aristotle as a Historian of Philosophy: Some Preliminaries". *The Journal for Hellenic Studies*, Vol. 77, No. 1: 35-41.

Guthrie, K. C. William. 1962. *A History of Greek Philosophy* (Vol. 1). Cambridge: Cambridge University Press.

Hankinson, J. Robert. 1998. *Cause and Explanation in Ancient Greek Thought*. Oxford: Oxford University Press.

Kahn, Charles. 1960. *Anaximander and the Origis of Greek Cosmology*. New York: Columbia University Press.

Liddell, G. Henry; Scott, Robert; Jones, S. Henry et al. 1996. *A Greek-English Lexicon*. Oxford: Oxford University Press.

Lloyd, C. Antony. 1976. "The Principle that the Cause is Greater than its Effect". *Phronesis*, Vol. 21, No. 2: 146-156.

Makin, Stephen. 1990-1991. "An Ancient Principle about Causation". *Proceedings of the Aristotelian Society*, n.s. 91: 135-152.

Mansfeld, Jaap. 1988. "De Melisso Xenophane Gorgia: Pyrrhonizinng Aristotelianizm." *Rheinisches Museum*, Vol. 131, No. 3-4: 239-276.

Marmodoro, Anna. 2017. "Gregory of Nyssa on the Creation of the World". In *Divine Powers in Late Antiquity*, edited by Anna Marmodoro and Irini-Fotini Viltanioti, 94-110. Oxford: Oxford University Press.

Mourelatos, P. D. Alexander. 1981. "Pre-Socratic Origins of the Principle that There Are No Origins from Nothing". *The Journal of Philosophy*, Vol. 78, No. 11: 649-665.

Mourelatos, P. D. Alexander. 1984. "Aristotle's Rationalist Account of Qualitative Interaction". *Phronesis*, Vol. 29, No. 1: 1-16.

Plato. 1929. *Timaeus. Critias. Cleitophon. Menexenus. Epistles.* Cambridge, MA: Harvard University Press.

Simplicius. 1882. *In Aristotelis Physicorum commentaria*. Edited by Hermann Diels. Berlin: Reimer.

Stannard, Jerry. 1965. "The Presocratic Origin of Explanatory Method". *The Philosophical Quarterly*, Vol. 15, No. 60: 193-206.

Stebbing, L. Susan. 1933. *A Modern Introduction to Logic* (2nd ed.). London: Methuen.

Stevenson, G. John. 1974. "Aristotle as Historian of Philosophy". *The Journal for Hellenic Studies*, Vol. 94: 138-143.

CHAPTER ELEVEN

CO-CREATION: THE RELEVANCE OF HERACLITUS AND EMPEDOCLES FOR THE RENEWAL OF WESTERN PHILOSOPHY

PREDRAG CICOVACKI

I. The Present Crisis of Philosophical Thinking

There have been three major motivational forces in the history of Western philosophy. They can each be represented by a single word: wonder, doubt, and disorientation. The first definitive determination of philosophy was that it is motivated by the sense of wonder. As Plato and Aristotle both argued, philosophy begins with our wonder about the complexity and beauty of the world; such wonder prompts us to reflect about the world in a rational and systematic way. The main questions of philosophical thinking triggered by wonder are the following three: What is the true nature of the world in which we live? What is the true nature of humanity? What should be our relationship with the world in which we find ourselves?

For almost two thousand years, the sense of wonder pervaded the Western philosophical thinking and motivated it to tackle ever more complex issues and problems. By the end of the medieval era, the gap that had opened between the entrenched way of thinking and new discoveries about the world led to a profound sense of doubt. The beliefs, accepted for as long as people could remember, were beginning to break down. New advances, from those related to expeditions into unknown lands, to inventions of new instruments (e.g. telescope, microscope), and the developments of new theoretical constructions (like Copernicus' theory of the heliocentric system) undermined any sense of certainty, especially with regard to the first and most fundamental issue, the one regarding the true nature of the world.

Descartes managed to capture this spirit of doubt in the most radical way. He also succeeded in redirecting the focus of Western philosophical thinking from the first to the second fundamental question: from world to humanity; from object to subject; from the outside to the inside. The ultimate certainty is not external but internal: I think, therefore I am. The second question, which deals with the nature of humanity—especially with regard to its cognitive capacities and ethical norms—, became the central preoccupation of the modern era.

Modernity was a time of rapid developments. In this second major philosophical epoch, it did not take two thousand years to realize the limitations of this new approach. Despite many remarkable insights of Descartes and his followers, philosophy had failed to discover any reliable anchor in the subject, just as it failed to find it in the object in the previous epoch. Neither the rationalist orientation nor its empiricist counterpart, nor any attempt at their combination managed to reassure our doubts as to the true nature of cognition. Even science, with its rigorous approach, was unable to deliver anything more definitive than the theories sufficiently corroborated by evidence—until new evidence and theories undermine the foundations of the accepted views and remind us that they are but our attempts at the best possible explanations. We have to admit that all our knowledge is a human, an all too human affair, without at the same time implying that it is in all cases relative and arbitrary.

The modern turn toward the subject did not have a happier outcome with regard to its ethical reflection than it did with regard to the nature of cognition. Neither utilitarianism nor Kantianism succeeded in providing a convincing normative theory of human behaviour. Much less did they illuminate the true nature of our moral aspirations to us. Along with giving up on our highest moral aspirations, we have become increasingly satisfied with the pursuit of pleasure and the temporary fulfilment of our desires.

The third question, that of our proper relationship with the world, turned out to be no less complex, and perhaps even more disorienting, than the other two fundamental philosophical concerns. Perplexed and frustrated, we ended up with a "culture" so profoundly disoriented that our central motivation in this age seems to be capturable by only one motive: escape. We have become so disillusioned with any grand project of Western civilization, and so disoriented in our thinking (and feeling), that our only defence seems to consist in an escape from reality, which also includes escape from ourselves. To make matters worse, our current technological development—perhaps the only thing that is blooming in our disoriented post-modern era—offers us plenty of opportunities to lose ourselves in virtual reality.

Although an escape from reality can offer us a sense of relief, this sense can be, at best, only temporary. Nevertheless, such an escape can allow us some time to re-group, so that we can confront our situation with a renewed seriousness and a reawakened sense of purpose. More fruitful than any such escape is an attempt to retrace our steps, rethink our history, and re-examine our roots. When it comes to philosophical thinking, our roots are in the Presocratic tradition of the ancient Mediterranean cultures. Among the numerous thinkers working during that early period of Western thought (6th-4th century BC), there are two who deserve our special attention: Heraclitus and Empedocles. They are worth reconsidering, because they defended an especially dynamic and fluid conception of reality. Since such fluid and insufficiently developed positions, based more on vivid imagination than on discursive thought, permit us ample space for reinterpretation and new exploration, I suggest that we return to them and see whether we can glean new insights from these forefathers of Western philosophical thought.

II. Heraclitus and Empedocles

Heraclitus of Ephesus flourished around the year 500 BC. There are few well-attested facts about his life and many more legends. His preserved fragments leave us with several interesting ideas—more hinted at than outlined—that deserve our attention. One of them deals with the concept of *logos*. According to Heraclitus, the whole world is intelligent and alive. Thus, intelligence is not something detached from the processes of the world and enclosed within our mental categories and frames of thinking. Rather, intelligence suffuses the world—it permeates being in its entirety—and is itself something living and vibrating with vital force. While Heraclitus mused about the coincidence—and even the identity—of opposites, this did not keep him from teaching their conflicts as well. Despite Aristotle's proclamation that this view violates the principle of non-contradiction, we remain intrigued by Heraclitus' pronouncement that "War is the father of all things". Heraclitus taught the flux of all things, yet also their underlying unity. The apparent contradiction of such a view did not seem to concern him, and we should rethink why contradictions bother us so much. Heraclitus also postulated a fundamental analogy between the cosmic fire, permeating the entire universe, and our being. While Aristotle wanted to classify his viewpoint as monism, we should not rush in this direction. The cosmic fire and the logos of which Heraclites spoke must have something in common. It is left to our imagination, however, to supply the missing links between these ideas and to clarify both what their common elements might be and to what extent they overlap.

Thinking in metaphors and images, more than in concepts and reflective ideas, Heraclitus understood the nature of the human soul in terms of cosmic fire. He suspected that its true nature is to be "hot", which we can understand in a mundane way as being driven by emotions and passions. Like the Indian and Chinese sages in the same era—which Jaspers called the "Axial age"—Heraclitus proposed a cyclical (rather than linear) conception of the unfolding of natural processes, including the recurring existence of the universe and of all being. Quite in harmony with the paradoxical nature of his thinking, Heraclitus holds that the world is subject to recurrent cycles of destruction and renewal.

We know a bit more about Heraclitus's somewhat younger contemporary, Empedocles (c. 494-434 BC). Nevertheless, the life of Empedocles, and especially his death, also remain hidden behind a veil of mystery. Of the fragments that remained from Empedocles' philosophical poem, "On Nature", three impress me the most. The first of them expresses his idea of the multiple elements of which things may consist. Unlike his Ionian predecessors, who tried to find one ultimate *"arche"* of all things, Empedocles opened a way towards genuine plurality in our understanding of the world. Specifically, he spoke about four elements: earth, fire, water, and air. (This is comparable to the ancient Chinese thought, in which the fifth element, wood, was also included.) Even more impressive was Empedocles' conviction that the processes in the world are driven by two mutually opposing forces, which nevertheless complement and complete each other. He simply called them "love" and "strife", and understood them as an attractive and creative force (love), and as a repulsive and destructive force (strife). Even in our "sophisticated" times, we still explain many phenomena in terms of the conflicting forces of attraction and repulsion. The most interesting and consequential thought of Empedocles, however, was the one in which he postulated unity in plurality and plurality in unity. Although Aristotle would again raise his brows and complain of contradiction, we may need a thought like that for any healthy approach to our understanding of the world, of our own nature, and of our relation to the world. As love and strife need each other, unity and plurality presuppose each other just as fully. Empedocles may have intuitively hit upon the most promising way of thinking of the complexity of reality and of our place and role in it.

III. Reactions to Heraclitus and Empedocles

Heraclitus spoke about the eternally burning fire and the logos that governs the universe in such a way that "no one can step in the same river twice." Although most of his fragments refer to the cosmic order, I believe that he

was just as interested in the human condition, in the soul as our true self. Heraclitus came to believe that the pattern of the cosmic order and the pattern of human life were—essentially, although not in all details—one and the same: the cosmic fire burns in us as well, for we are also an illustration of the paradoxical plurality in unity and unity in plurality. Especially when recalling Parmenides' denial of motion and change and the later Western preoccupation with being in opposition to becoming, I am in awe of Heraclitus' views. They open a way not only for a dynamic conception of reality, but also for that which Leibniz later called "perennial philosophy".

Not trailing much behind the prophetic voice of Heraclitus, Empedocles brilliantly postulated that there are two governing principles of the cosmos: love and strife. It is important that, for Empedocles, they were not reducible to good and evil. All we can claim is that they were conflicting yet complimenting. Because of these two clashing forces, the universe had to be understood in dynamic rather than static terms: as eternally becoming rather than as complete and frozen in its being. This is what permits Empedocles to maintain that the cosmos is equally a unity in plurality and a plurality in unity.

The more we think about these fascinating views, the more we realize that the dominant current in subsequent history of Western philosophy failed to develop these insights. It is even more accurate to say that it attempted to reverse them. What is so fundamentally unsettling about their dynamic view of reality is that it does not promise us any sense of closure. We would like, for example, the meaning of life to be something predetermined, something woven into the fabric of reality. We hope for a life that is so well-structured that it can approach perfection. Yet, despite our stubborn insistence, we cannot achieve any such closure. Life offers no such promises and no protection against sudden, and often unjust, turns of fate; it puts forward no view that can be hardened into a never-changing and forever-valid dogma.

Interestingly enough, the explicit criticisms of Heraclitus and Empedocles do not directly mention our deeply ingrained fear of life and of a world without closure. Although such concerns lurk in the background, the criticisms go after two related but distinguishable targets. The first complains about the lack of any final purpose in the universe in the views of Heraclitus and Empedocles, and the second protests that their views violate the laws of logic.

The Greeks widely believed that there had to be some ultimate, divinely conceived design that would justify our existence in the world. Is not the universe, and our respective lives within it, guided by an irrevocable

destiny? How, then, could Heraclitus and Empedocles not rely on, and subsequently postulate, the existence of such divine elements?

Such criticisms came from both the later Greek and Christian sources. Among Greek philosophers, it was hardly possible to find those who maintained that the cosmos had no inherent meaning whatsoever. Only centuries later, with the collapse of the second major epoch of philosophy, when the religious way of thinking fell into disrepute and modernity began to slip toward relativism and nihilism, the view asserting that the world has no intrinsic meaning or purpose became accepted as the mainstream position.

Heraclitus and Empedocles' way of thinking allows, however, for a third possibility: the world has no meaning predetermined for it from the outside, but this does not imply that the world has no meaning whatsoever. The denial of divinely preordained meaning leaves enough room for some meaning to be created and introduced from the inside, in the course of its dynamic process. This meaning is not final and static but has to be changing and adjusting from within. As Heraclitus expressed it, our character is our destiny.

The second point of attack concentrates on the view that this offends the logic of our rational thinking by postulating that the world embodies a unity in plurality, and a plurality in unity. The criticism is that this simply cannot be the case, for it would destroy the internal coherence of our rational thinking. In order for us to conceive of the world in a rational way, it has to be either unity, or plurality. There is no third way.

Wallace Stevens, the famous American poet, caricatured this logic of either-or as the click-clack of reason. This logic, together with its "laws" that are allegedly the laws of all thinking (and even of all being), lacks not only any trace of humour but is also detached from the flow of real life. It is detached from life that is full of paradoxes and antinomies, wonders and miracles, which the click-clack of reason can neither explain nor explain away. What may look like incompatible alternatives to our reason, the flow of life presents as mere alterations and endless transformations. Behind the authority of Aristotle, this click-clack rationality has become so entrenched that it has affected virtually all aspects of life. While poets and artists in general have, in the last two centuries, continued to experiment with new forms of thought and expression, it may be that a fundamental change in paradigm has to come from those who have "fixed" the previous forms of thought and expression. If Richard Wagner is right to claim, in his opera *Parsifal*, that the wound can heal only the weapon that makes it, the healing of the wound of our disorientation would have to come from the source that has made it: from our established way of philosophical thinking.

IV. Resisting the Temptation to "Fix" Things

Perhaps the source of our deepest wound is in our desire to "fix" things. We have an urge to fix what is dynamic and fluid, ambiguous and open-ended. Us fixing things may stabilize and disambiguate them, but we have not been sufficiently aware that these practices come at a price. Nor have we realized that this price may be too high. What exactly is the price and what makes it so high?

If something is by its nature ambiguous, some of its dimensions and aspects must be eliminated or distorted in order to make its boundaries clear and definitive. That in turn may mean coercing things to become what they are not and forcefully changing their nature, so that they fit into our pre-designed categories and classifications. Put more bluntly, by "fixing" things we damage and mutilate them.

What we get at the end of such a process is similar to what happens after we "educate" our children: we succeed in reducing those curious and rebellious little creatures into obedient and efficient members of our society. They stop asking questions like: And who made God? Instead, they become more concerned about obeying the norms of behaviour that we drill into them. If we are not completely damaged and mutilated ourselves, we have to wonder what might have happened if we ceased to insist on imposing the same uniform standards of behaviour on all children and if we tried to be more sensitive to their individual gifts and dispositions. As thinkers, we have to ask the same questions with respect to the nature of the world, our own nature, and the nature of our proper relationship with the world.

Heraclitus and Empedocles pointed to the possibility of a world without closure, a world in which things partially overlap and are both one and many. If we take these thoughts seriously, we must entertain the possibility that the world has no fixed beginning and no fixed end. At first glance, this may seem confusing, because in the Western world we have become conditioned to think that the world must have a beginning and an end. Not only our religious tradition but even our science stubbornly insists on the creation myth, with respect to both time and space. Think, for instance, about the Big Bang or the moment of an explosion and expansion of an infinitely small amount of matter. We consider that moment as the beginning of natural history, and have designed numerous experiments that reconstruct the exact temporal sequence following the explosion. While doing that, we are apparently forgetting that what allegedly exploded and thereby began its expansion is an infinitely small amount of matter of which we have no clue where it came from. And without that "little detail", our story cannot be the story of the beginning, and our creation myth is as

unsustainable as it has ever been. At best, it could be a story not of one creative force acting at a definitive time (and space), but a story of continual creation, and creation by multiple forces, as Heraclitus and Empedocles envisioned so many centuries ago.

Perhaps the most important thing we can learn from Heraclitus and Empedocles is to resist the urge to force all our thinking into "either-or" categories. They indicate a pattern of thinking that can be called "both … and". In the world, there is *both* love *and* strife. The world is both a unity in plurality and a plurality in unity. There is a continuous flux of things, yet they remain constant as well.

The acceptance of this way of thinking allowed Stéphane Lupasco to make a significant break from the traditional logic of the excluded middle. Although numerous experiments in quantum physics have undoubtedly indicated simultaneous presence of seemingly contradictory elements (such as wave and corpuscle, continuity and discontinuity, reversibility and irreversibility of time), these pairs appear mutually contradictory when they are analysed in terms of classical logic. This logic is founded on three axioms: (i) The axiom of identity: A is A; (ii) the axiom of non-contradiction: A is not non-A; and (iii) the axiom of the excluded middle: there exists no third term T which is at the same time A and non-A.

Lupasco developed the logic of the *included* middle: there exists a third term T that is at the same time A and non-A. He resolved the problems of contradictory pairs by using the idea of multi-layered reality (see Lupasco 1987). As Basarab Nicolescu sums up Lupasco's discovery, "The third dynamic, that of the T-state, is exercised at another level of Reality, where that which appears to be disunited (wave or corpuscle) is in fact united (*quanton*), and that which appears contradictory is perceived as noncontradictory" (Nicolescu 2002: 29; see Brenner 2010: 248-271).

Although Lupasco's new logic has had a powerful (albeit underground) impact among psychologists, sociologists, artists, and historians of religion, it has been marginalized by philosophers (and physicists). It may be that philosophers have been too preoccupied with their own objections to classical logic and their own attempts to develop a viable alternative to it. For example, the same principle of the excluded middle, together with the principle of bivalence ("Every statement is either true or false"), has recently come under attack by the prominent philosopher of language, Michael Dummett. The motivation for this attack should sound familiar: the dynamic flow of many processes in reality makes it difficult, if not impossible, to establish whether many of our statements are either determinately true or false (see Dummett 1991, 9-10, 17, 74-75).

One of the reasons why it is so difficult to resist the temptation of the "either-or" thinking is that the pattern of "both … and" makes it seem that there is much in it that is irrational. What, however, is wrong with things being irrational? And, conversely, why does being rational appear so intrinsically valuable?

As human beings, we are both rational and irrational. Our initial contact with the world is not based on rational ideas but always on our emotional experiences. It is through such experiences that we come to feel both the "hardness" of that world and the resistance of the real, which serves as an invaluable check against our beliefs and ideas about it. The world must be permeated by logos, as Heraclitus postulated, and so must be our thinking about it. But must this logos be understood as detached from emotions and imagination, as Aristotle and his followers have succeeded in convincing us?

Another part of the problem with our exclusive "either-or" reasoning can be seen in the following example. For any two X and Y, our reasoning leads us to believe that they must be either identical, or that they are not so. Reality, however, is not nearly as inflexible and unambiguous as our thoughts. Reality is full of cases that can be classified as "partial identity", which, moreover, can increase or decrease with time and the flux of events. Let us illustrate this on one example that would be closer in time to both Heraclitus and Empedocles. In *Symposium*, Plato relates, through Aristophanes, a fable according to which human beings were originally far different than they are now. In fact, Plato's Aristophanes claims that people are now only half of what they used to be. We were all full circles and, out of jealousy, the gods divided us into halves. In his eulogy to Eros, Aristophanes argues that love consists of becoming united with our missing halves; when we find our missing halves, we feel complete again. A more sophisticated version of this symbolism was advocated by Carl Gustav Jung and his idea of a closed and complete circle as a symbol of fulfilment and perfection. Relying on numerous mystical and religious traditions, Jung called it by the Sanskrit word "*mandala*."

The problem with the circle-symbolism is that there are no complete circles in nature, just as there is no closure in reality. The ending of one cycle is at the same time a beginning of the new one. The relationships of people, and of many interconnected aspects of reality, is much better represented by the symbolism of the partially overlapping circles than by one full circle.

The area of the partial overlap of two circles looks like an eye, or like an almond nut. "*Mandorla*" is an Italian word for almond, and sometimes this word is used both for the area of the partial overlap and for the entire figure.

It is confusing, however, to use the same word for two such different things. This is why I suggest keeping the word "mandorla" for the overlapping (and almond-shaped) area and introducing another word, "*pandorla*" (pan=all), for the complete figure of the overlapping circles.

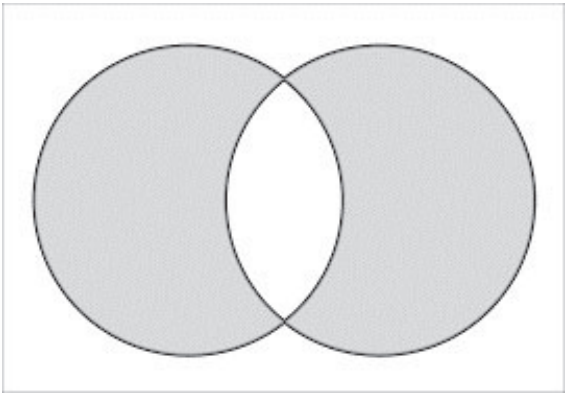

Figure 11-1: A simple pandorla

In his book, *Owning Your Own Shadow*, Jungian psychologist Robert Johnson uses the term "mandorla" in the double sense mentioned above. More importantly for our context, he makes a contribution toward a more adequate understanding of our (philosophical) thinking by making a significant distinction between paradoxes and contradictions. Johnson argues that the idea of partial identity, or of the partially overlapping circles, is difficult to accept because it looks like a contradiction: either X and Y are identical, or they are not. He points out that such ideas may be paradoxical, but they are not contradictory. Johnson insists that this difference between paradoxes and contradictions is not only clear but also of enormous significance for how we understand our world, ourselves, and our relationship with the world.

In Johnson's view, contradictions are unlike paradoxes insofar as they impose the crushing burden of meaninglessness on us. Following Nietzsche, Johnson points out that we can endure any suffering if it has meaning, but meaninglessness is psychologically unbearable. Contradictions are barren and destructive, while paradoxes are creative: they represent "a powerful embracing of reality". This is one of the reasons why all religious traditions, Christianity included, express their teachings in terms of paradoxes; while contradictions are static and unproductive, paradoxes make room for grace and mystery (see Johnson 1991, 74-75). To illustrate his idea, Johnson

reminds us of the words from the Gospel of John (12:24), "Verily, verily, I say unto you except a corn of wheat fall into ground and die, it abideth alone: but if it die, it bringeth forth much fruit."

Johnson applies these insights regarding the partially interconnected nature of things and the pandorla symbolism to our language as well. While we may not (yet) have adequate names and concepts for the pandorla symbolism, our language itself is an illustration of the pandorla-interconnectedness of words. According to Johnson, every well-structured sentence is a pandorla. Our principal verb *to be* is a great unifier. A sentence with the verb *to be* is a statement of (partial) identity which bridges the split between two elements. This is confirmed by observing the subjective form for the predicate of the verb *to be*: we say, "I am he", not "I am him". This claim implies that he and I are the same, which is a statement of the paradoxical unity in plurality.

Johnson maintains that all sentences make similar identity-claims even apart from the verb *to be*, though it may be less obvious. To make any well-formed sentence is to make unity out of duality. One makes a pandorla every time one says something.

Every time Johnson uses the word "identity", he should be taken to mean *partial* identity. In some cases, instead of identity, it would be more precise to talk about relatedness and interaction, or overlap and intersection. He should also be reminded that every sentence retains a duality—more precisely, and just as Heraclitus and Empedocles saw it: a multiplicity—while being a unity as well.

More important than these small corrections are the following insights. Once we have the pandorla symbolism in place, we can see both how well it captures some of the ideas of Heraclitus and Empedocles, and also how much more adequately it represents human relations than our usual either-or way of thinking. Here is a useful way to illustrate it. With those we do not know, there would normally be no intersecting area. With those we barely know, the circumferences of the circles may touch but need not overlap. The less intimate the relationship, the less overlap there will be. Conversely, the better we know someone and the closer we are, the more extensive the overlap. Such relations are not set in stone; they have to be re-created and continuously cultivated by all of those whom they involve.

The pandorla symbolism enables us to explain yet another common and important phenomenon. With regard to any one of us, there are numerous, virtually countless numbers of overlaps with other beings: humans, animals, plants, the environment, works of art and other cultural and historical entities, as well as inanimate objects. As Heraclitus and Empedocles clearly saw, all of these relationships are dynamic: there can be periods when some

people (and things) mean little to us and other times when they are very important.

Keeping those points in mind, we should realize that the number of overlapping circles is hardly ever just two. Each one of us, as well as every single entity that exists, is involved in a virtually unlimited number of pandorlas. Not surprisingly, then, there is an endless number of combinations that can be made with the overlapping pandorlas. Because of its historical significance, I will here call attention to the one that has traditionally been called the "flower of life". This is a hexagonal pattern where the centre of each circle is in the circumference of six surrounding circles of the same diameter, which is made of nineteen complete circles and thirty-six partial circular areas. Although life and the relationships that we find in reality are never so regular, and there is no one circle encompassing it all—except perhaps for our concept of the world or the universe—this dynamic pattern can be visually represented as follows.

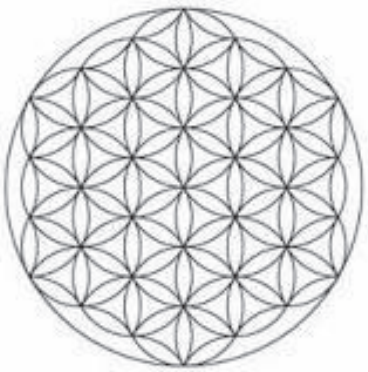

Figure 11-2: The flower of life pandorla

This symbol is so archaic that it can be found in some of the earliest civilizations. The oldest known representation of the flower of life is from Egypt, and it is at least six thousand years old. In the Temple of Osiris at Abydos, this figure is burned into marble, rather than chiselled into it. In India, this symbolic representation of life exists in the Golden Temple in Amritsar. We can also find it in the Forbidden City in China. This archaic symbol illustrates a unity in plurality and a plurality in unity, as well as the overlap and inter-dependence of existing things.

V. Hartmann's Conception of Reality

Nicolai Hartmann is one of those rare philosophers who use the symbolism of pandorla (see Hartmann 1954, 87, 93, 105; Hartmann 1965, 365). He also reconstructs the nature of reality and our role and place in it in the dynamic and fluid terms that very much resemble Heraclitus and Empedocles. According to Hartmann, the world is without closure, without beginning and end; all our encounters with it teach us that it is a genuinely pluralistic and dynamic world. Despite the endless philosophical debates about monism and dualism, there are four discernible and mutually irreducible layers of the real world: the inorganic, the organic, the psychic, and the spiritual. Hartmann postulates three dynamic laws regulating the relationships of these strata. The law of strength says that the higher and weaker layer of existence is conditioned by and dependent on the immediately lower and stronger layer. The law of material claims that every lower type of existence is merely material for the higher type that is raised on it. The law of freedom asserts that, compared to the lower stratum, every higher layer of real being is an entirely new structure and that it is free, as over against the lower type.

Hartmann's ontological scheme of the four layers of real being may be represented by the following pandorla:

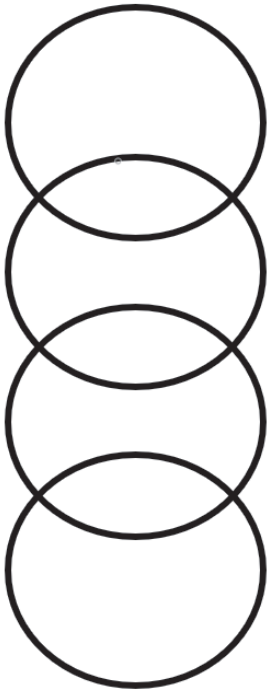

Figure 11-3: Hartmann's pandorla of the four strata of real being: the inorganic, the organic, the psychic, and the spiritual

As far as we know, there are only these four layers of reality; they are the co-creators of the real world as we know it. Hartmann emphasizes, however, that this does not mean that the four layers have always existed, or that in the future new layers will not be discovered or developed. There is nothing about the nature of reality that fixes their number to four, nor to exactly the four with which we are familiar and to no others.

In two of his ontological works—*Der Aufbau der realen Welt* and *Neue Wege der Ontologie*—Hartmann distinguishes the categories specific to each stratum of reality. As the categories of the corporeal world he identifies: space and time, process and condition, substantiality, causality, and reciprocity, as well as dynamic structure and dynamic equilibrium. The categories of animate nature include: adaptation and purposiveness, metabolism, self-regulation and self-restoration, the life of the species, the constancy of the species and variations. The categories of psychic reality involve: act and content, consciousness and unconsciousness, pleasure and displeasure.

Finally, the categories of the spirit are: thought, knowledge, will, freedom, judgment, evaluation, and personality. There are no dominant categories within a single stratum, but they all jointly determine everything. As a result, it is impossible to grasp any single category by itself.

In addition to the categories specific to each stratum, Hartmann detects some categories that run through the entire sequence of strata, although in varying forms. Such categories are: unity and multiplicity, concord and discord, discretion and continuity, substratum and relation, element and structure, form and matter, inner and outer, determination and dependence, identity and difference, generality and individuality, as well as the modal categories and their negative counterparts (see Hartmann 1949, 157-374; Hartmann 2012, 59-101).

The most unifying theme in Hartmann's pluralistic worldview is that of opposition and dynamic balance. Opposition is not to be confused with contradiction, which Hartmann believes exists only in thought. Every known structure in the real world, from atoms to solar systems, from the simplest living organisms to human beings, displays a complicated array of counter-forces as well as continuous attempts to balance them. For Hartmann, there is no independence without dependence. More precisely, all there is is an inter-dependence: both partial independence and partial dependence, and they complement each other very well. For instance, there are two ways in which the higher mode of being is dependent on the lower: the first is existential (e.g. spirit cannot exist without a supporting consciousness and, indirectly, a body), and the second is limiting in terms of content and structure (e.g. the lower mode of being provides matter and serves as a basis for reshaping and rebuilding the higher form of being). These two forms of dependence can also be used to illustrate the basic laws regulating the mutual relationship of different strata of reality. They are the law of recurrence, which guarantees partial continuity between the various strata, and the law of novelty, which ensures diversity. The determining power of matter does not extend beyond its limiting function; it does not prevent the novelty of the higher form but rather merely limits its scope. The real world is not governed either by matter or by spirit. It can be ruled neither from below nor from above, for its every layer, besides continuity, includes a certain irreducible specificity. Thus, the real world is an intricate, perplexing, multi-layered, and dynamic unity in heterogeneity; and insofar as we are complex and multi-layered beings in the real world, so too are our relations with this multi-layered reality.

VI. Concluding Remarks

While unorthodox and not sufficiently pursued by later Western thinkers, the ideas of Heraclitus and Empedocles may have themselves been the illustration of an even older way of thinking. Such thinking may have been considered too dynamic by the vast majority of later philosophers: it was too flexible to categorize into a stable worldview, because it calls for continuous creation, re-creation, and co-creation of the world, of ourselves, and of our relations to the world. The reaction of later philosophers to the poetic and imaginative thinking of Heraclitus and Empedocles was to try to fix things by delineating the firm boundaries of the world and by establishing the corresponding laws of thinking.

We have certainly gained much by turning away from thinking in terms of continuous co-creation in the manner of these Presocratic thinkers. But we may have lost even more. We have managed to disambiguate the world and to force its processes into stable and predictable patterns that have served us for more than two thousand years. Yet these patterns have begun to feel like chains. In the natural sciences and arts, serious attempts have been made to rethink the old patterns and to provide us with alternative ways of experiencing the world and our place and role in it. In retrospect, philosophers have turned out to be far more conservative than both scientists and artists. Has the time come for philosophers to join the efforts of these innovators in trying to develop a new, fourth paradigm of thinking about reality and of orienting ourselves in it?

If we would like to sum up this new paradigmatic orientation in one word, it could be “co-creation”: the world creates us, and we in turn create that world and ourselves in it in a never-ending process. In trying to develop this new way of philosophical thinking about the nature of the world, our own nature, and our relationship with the world, we will find much to learn and much to inspire us by revisiting the two sages of the old, Heraclitus and Empedocles.

Bibliography

Brenner, Joseph E. (2010). "The Philosophical Logic of Stéphane Lupasco (1900–1988)". *Logic and Logical Philosophy*, Vol. 19: 243-285.

Dummett, Michael (1991). *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.

Hartmann, Nicolai (1949). *Der Aufbau der realen Welt*. Berlin: Walter de Gruyter.

Hartmann, Nicolai (1954). *Einführung in die Philosophie* (3rd Edition). Tübingen: Karl Auerbach.

Hartmann, Nicolai (1965). *Gundzüge einer Metaphysik der Erkenntnis*, (5th Edition). Berlin: Walter de Gruyter.

Hartmann, Nicolai (2012). *New Ways of Ontology*, translated by Reinhard C. Kuhn. New Brunswick, NJ: Transaction Publishers.

Johnson, Robert A. (1991). *Owning Your Own Shadow*. San Francisco: Harper.

Lupasco, Stéphane (1987). *Le principe d'antagonisme et la logique de l'énergie* (2nd Edition). Paris: Éditions du Rocher.

Nicolescu, Basarab (2002). *Manifesto of Transdisciplinarity*. Trans. Karen-Claire Voss. Albany, NY: SUNY Press.

CHAPTER TWELVE

ARISTOTLE AGAINST ACTUAL INFINITE REGRESS

MATTEO COSCI

Introduction: Aristotelian Infinite Regresses as Thought Experiments

Infinity is a concept that is not as easy to grasp as it may seem at first glance. Perhaps we may ignore it on purpose, entirely avoiding dealing with such an immense and indomitable subject. However, it has been suggested that "our minds seem to *require* the idea that things might 'go on forever'—in space and time, in the future and in the past. Infinity is perhaps, a mental default, a natural side effect of the pattern-seeking abilities of our minds" (Stewart 2007, 3), so that as long as we try to dismiss such a requirement, it may return to be accounted for. Nonetheless, infinity might also be a cognitive bias emerging from a structural or systematic error of how things are perceived or of how we presume they can go on. In this sense, it has been defined as a sort of unwarranted form of reasoning *by recurrence* (see Falk 1994) that may arise in thought to bypass dimensions that we *cannot* otherwise embrace in their entirety at once. If infinity thus originates *in our minds*, then it is an issue that cannot be ignored or easily expelled from within.

Originally, ancient Greeks had a negative conception of the infinite. What they felt was *"horror infiniti"*: it was an idea that scared their self-confidence, like the perspective of an ocean that cannot be traversed to reach a safe shore. For the ancient Geek mind-set, even the hint of infinity was basically a sign of uncanny vagueness to be avoided to preserve their ideals of order and completion. The word "infinite" (*ἄπειρον*) literally meant something between *indefinite* and *limitless*: an unbounded and dangerous extension. For them, generally speaking, what was constitutively unfinished was usually associated with imperfection (see Mondolfo 1956).

Aristotle, one of the cleverest and most influential thinkers on the subject, had an idea of *concluded* perfection, especially in reference to what is fully-developed and entirely self-contained, i.e. "that have no parts outside itself" (*Metaph.* Δ, 19, 1021b 30-1022a 1), which was *in contrast* to his definition of infinity as "that of which there is always something outside" (*Phys.* III, 5, 204a 20). Underlying Aristotle's texts is the idea according to which infinity, in so far as it is incomplete and thus unfinished and unaccomplished, represents imperfection, while, on the other hand, nature (*φύσις*) is the principle of order, determination and purposefulness, which even "abhors" and "finds repulsive" the infinite.

The reason for this latent contraposition between such a completed perfection and a constitutively uncompleted infinity is to be sought in the presence or absence of a *limit*. For, having a limit always meant for Aristotle having some aim, because "limit" (*πέρας*) consists also in "the *end-aim* (*τὸ τέλος*) of each thing" (*Metaph.* Δ, 20, 1022a 4-6). Therefore, infinity's limitlessness and unachievability is implicitly intended by him as a flaw of aimlessness or absence of scope. As such, infinity is something that stands against the idea of a purposeful nature and the finalized way, it seemed to him, in which it always works. According to these premises, then, for him there is no space for the infinite in an orderly and settled cosmos. Such an exclusion should have been maintained especially if one thought, as Aristotle did, that the world is *all enclosed* in its furthest spherical boundary and nothing more exists beyond that extreme limit.

Surprisingly enough, however, Aristotle argued that in one sense infinity indeed exists, but not in another. He famously distinguished between actual infinity and potential infinity, maintaining the former impossible under every aspect, while granting only some subsistence to the latter, e.g. in the infinite divisibility of geometrical extensions. Even so, the peculiarity of being-in-potency of such an infinite lies for him in the fact that it can never become actual, i.e. it can never reach its full development of the entirety of all its parts at once. So, differently from how he usually intended standard potentiality (as for instance the seed being the tree in potency), the potential infinite can be actualized only as part, or phase, of a process underway. The instance of the infinity that now-and-now comes to actuality, Aristotle specifies, is always something diverse and finite. Not even time, which for him is eternal, is actual infinity, because its parts, or constitutive moments, do not come to existence together but only *one after another* in the present. Thus, the fact that time can be infinitely divisible does not imply that it is infinitely *divided*: that would impede chronological succession, as is not the case.

Considerations about how infinity as such was intended by Aristotle can be found in the classical pages of his *Physics,* where he developed his own account on the topic (III, 4-8). However, many other pertinent considerations can be found in different passages of his other works as well. One way to recover some scattered references and at the same time to see how he intended to deal with infinity "in practice" is by looking at his so-called *infinite regress arguments*.

Infinite regress arguments are lines of reasoning that are based on a series of consequent elements, with some premises but *no conclusion*, because each member of the series replicates what it requires or entails to the following, and the following to the next in a never-ending succession of reiteration (called "infinite regress"), so that the actual end of the argument turns out to be infinitely postponed. Depending on the context, infinite regress arguments can be considered vicious or non-vicious. An infinite regress argument is usually considered vicious if it entails at same point either a false statement, a contradiction, or an unacceptable consequence, such as, for instance, if a proposed solution proves itself wrong, unresolving or inconsistent along its argumentative derivation. Aristotle's infinite regress arguments are often concluded by the expression "*εἰς ἄπειρον*", *to infinity*, which is for him a mark of bold inconsistency. Infinite regress arguments are usually aimed by Aristotle to perform a *reductio ad absurdum*, the perspective of an actual infinity being always impossible or inconsistent in respect of his assumptions. For Aristotle, *all* infinite regress arguments are vicious and should be rejected as such, because for him the occurrence of infinite regress is sufficient evidence that the theory under analysis has departed from reality and is therefore inconsistent with the finite nature it should describe.

Specifically, for Aristotle, infinite regress arguments are not vicious *as arguments*, since they perform their function of hunting down infinite regresses perfectly. What is vicious are their envisaged contents, i.e. *the infinite regresses* that such arguments entrain. In this sense, infinite regress arguments are rather beneficial for Aristotle's reasoning, as proved by the fact that he made frequent and consistent use of them in his philosophical works. By using them, all he wanted to show as inconsistent were those regresses to infinity that pretend to be actual or actualisable. Aristotle's infinite regress arguments are, in fact, intended to display how the contested infinite regresses, whenever they occur, occur *only in a potential way*, since for him it would be impossible otherwise. Accordingly, Aristotle's regress arguments are infinite regress arguments in a potential state too, because it is impossible that they come to actuality if their object is compelled to remain in a potential status that never actualizes in full. Or rather, the

actualization of an infinite regress argument can be attained only when the infinite regress is set down, exhibited as such, shown as inconsistent for some reason, and finally rejected. That is probably an infinite regress argument "in act" in the Aristotelian sense, whose "actualization" corresponds to relegating the pretended infinite regress to its not-otherwise-possible potential status, blocking its occurrence at its inception. As mentioned,

> the impossibility of an infinite regress—the idea that anything can only be and only be thought under a finitude of conditions—has been taken to constitute one of the most fundamental principles of Aristotle's philosophy. Such a rejection of infinite regression recurs repeatedly in the works of Aristotle on the basis of his denial of actual infinites, seeing that unending regresses can never be brought to completion in something that is actually infinite. Here his position is straightforward: whatever is real must be complete, a whole,—and an infinite series cannot be completed. (Rehsher 2010, 105)

Thus, a thesis that implies infinite regress is usually introduced by Aristotle to show the inconsistencies it brings about and is consequently to be disregarded. That thesis is something more and something less than a simple working hypothesis. It is something less because, from his point of view, it cannot have any reference to real and existing instances in the external world as one may expect. It is something more because, from his point of view, it is overly made up and fictitiously crafted by imagination for the sake of its rejection. In this sense, every Aristotelian infinite regress argument establishes an inventive and imaginative scenario that can exist only in thought insofar as actual infinite regresses are involved.

It is certainly to be granted that, from a formal point of view, "Aristotle's notion of infinite is not necessarily involved in establishing the viciousness of infinite regresses" and that "we can establish the viciousness of most infinite regresses in Aristotle's works without assuming that he tacitly uses the claim that actual infinities do not exist" (Gratton 1999). However, Aristotle very often inscribes his rejection within his framework of a spatially finite universe. In fact, as mentioned, Aristotle believed that the world we live in is bounded in an all-inclusive and all-encompassing sphere (see *Phys.* III, 5; *De Cael.* I, 5, 271b 1-7, 276a 17). This assumption always remains at work, even when it is not explicitly recalled, in the course of his objections to a pretended infinite regress "in act". For him, if one accepts infinite regresses as actual, especially those which have some spatial import, that entails admitting that the universe is somehow infinite. But an infinite universe for Aristotle means an *imperfect* universe. Besides being logically inconceivable, it would (contradictorily) mean to live in an unnatural nature

within a disordered cosmos. But Aristotle has to introduce made-up scenarios, which are impossible in his worldview, to show what kind of inconsistencies would arise if one accepted the presumed regresses as actual and valid. In this sense, it can be suggested that the infinite regress arguments that Aristotle addressed against the pretended existence of actual infinities can be considered as loose forms of thought experiments.

Even if—strictly speaking—the ancient Greeks had no conception of thought experiments as such and did not even have a word for "thought experiments", it nonetheless seems that

> Aristotle uses [what we are inclined to call] thought experiments for argumentative persuasion and in places where, due to the obscure nature of the subject matter or the counterintuitive nature of the thesis they are meant to support, insight cannot be readily communicated by appeal to observational facts. (Corcilius 2017, 73-74)

Since it is hard to have a concrete representation of what actual infinity might be like, Aristotle's infinite regress argument (or his "inventive argumentative strategy", as Corcilius puts it) forces us to stretch our minds to show that such a picture is even impossible.

Aristotle's "thought experiments" on infinite regress seem to, more often than not, respect the main condition that a modern thought experiment should satisfy to be defined as such, namely to envisage imagined and replicable scenarios under a certain degree of control of their conditions with the aim of setting otherwise unfeasible experimental trials for testing *in thought*, and thus providing some theoretical evidence on what could not be assessed by perceptual experience. Now, actual infinity is not something that can be experienced under any condition (see *Metaph.* α, 2, 994a, 24-26). What can be experienced is its impossibility and its cognitive unsuitability, since infinity, and in particular actual infinity, is not something that humans can assess by perceptual experience. Aristotle thought that actual infinity could be not assessed in any way because it does not exist at all. What could be assessed is potential infinity and it could be assessed by way of infinite regress arguments as falsificationist thought experiments addressed against the pretension of existence of actual infinities.

From an anthropological point of view, it also even seems that "is not the capacity of thinking infinite regress as such but rather *the ways that have been developed to interrupt or inhibit infinite regresses* that are what is really characteristic of the human life" (Virno 2010).

On the Necessity of Avoiding Infinite Regress according to Aristotle

In what follows, I take in consideration thirty Aristotelian arguments, subsumed under eight general headings (from *a* to *h*). The purpose is to offer a comprehensive overview of Aristotle's infinite regress arguments, broadly intended as falsificationist thought experiments, aimed to avoid actual infinity on many levels, from logic to metaphysics, ethics to politics, physics and physics of the elements, to psychology and biological works, cosmology to anti-Platonism. The common thread is avoiding infinite regress at all these levels. In conclusion I will present the last selected cases as a general reappraisal of Aristotle's peculiar elaboration of the topic.

a) On the necessity of avoiding infinite regress in predication, universally quantified propositions, definitions and principles seeking, syllogisms and chains of syllogisms

[1] *Terms.* Should everything be an accident, then all predicative relations were constituted by accidents. In that case, according to Aristotle, we would find ourselves unfounded in our attempt to say such-and-such of a certain subject, because we will lack the substantial grounds that are necessary to predicate something on something as its actual reference. In that scenario, a *regressus ad infinitum* will occur, since an accident could only be predicated on another accident, which in turn could only be predicated on another accident, and so on and on (*Metaph.* Γ 4, 1007a 33 - b 3). In the absence of any subject of reference, the operation of predication would be impossible, since an accident cannot be an accident of an infinite number of accidents (e.g. "capable-of-making-shoes" of "good" of F_n etc.). Moreover, for Aristotle, an accident can neither predicate, nor be predicated on itself (e.g. "white" on "white"), nor re-predicate something that has already been predicated on it (e.g. "white" on "white man") at the risk of potentially infinite redundancy (*De Int.* 11, 20b 31 - 21a 7). In fact, accident is by definition something that is always "said of other", so that it always requires some kind of referential basis to stand as such. Its reference needs to be something that "cannot be said of anything else" and that reference is *substance*. Therefore, for Aristotle, substance as subject of predication is what can *prima facie* prevent us from referential infinitism.

[2] Analogously, it would be a mistake to try to find separate definitions of *per se* attributes, since those attributions can be intended only in reference to their peculiar and exclusive subject (e.g. "snub-nosed" as a *per se* attribute of "nose" only), so that any attempt in defining them will result in

non-essential, circular explanations (e.g. "snub-nosed is what is said of that nose that is snub-nosed", *ad infinitum*). Such results cannot even be considered to be definitions at all (*Metaph.* Z 5, 1030 b 35).

[3] For Aristotle, definitions cannot be constituted of an infinite number of definitory terms, because an infinite list of definitory terms is not distinctive as a definition should be, i.e. on the essence of the defined subject, and therefore is too indeterminate to be accounted for in any scientific argument, as Aristotle expected. For this reason, scientific syllogisms cannot be based on definitions composed of a limitless number of definitory terms, nor can they have an infinite number of intermediate *per se* terms among their premises, as they will be inconclusive otherwise. Furthermore, in the case of deductive chains of scientific syllogisms, the predicative series on which such syllogisms are based on cannot be infinite in number, neither in their descending line, nor in their ascending line, otherwise the deductive chain will be unfounded or inconclusively endless (*Post. An.* I, 19-22) so that it cannot be verified as true. For, according to a principle that has been aptly called "Aristotle's razor", or principle of economy in removing infinity from theory and from nature, "not only is explanation in terms of the finite always easier and better than in terms of the infinite, but infinity in all respects should be a suspect concept and avoided even at the cost of making explanations more complex" (Drozdek 2008, 102).

[4] *Propositions.* For Aristotle, the Heraclitean claim according to which "all claims are true" is self-refuting and thus must be rejected because it itself states that the contradictory claim according to which "all claims are false" is also true. Therefore, if one would say that "all claims are true" *except* the claim "all claims are false", an infinite regress would arise, because if the claim "all claims are true" is true then the claim that "it is true that all the claims are true" will also be true and so on. Therefore, such a statement is found to be unfounded and inconsistent. (And the same happens if one starts with the converse Protagorean claim "all claims are false", making the exception for the claim according to which "all claims are true"). Here Aristotle makes use of infinite regress arguments to refute universally quantified statements (such as the aforementioned which, as shown, "collapse themselves", *αὐτοὺς ἑαυτοὺς ἀναιρεῖν*) to demonstrate their existentially quantified counterparts, namely that rather "there is (at least) something that is true" and "there is (at least) something that is false". In the context of the proof of the excluded-middle principle, this logical operation serves his purpose of excluding that further possibilities are given, opposing at once radical relativism and scepticism alike (*Metaph.* Γ 8, 1012b 33 - b 3).

[5] *Syllogisms.* Within the Aristotelian architecture of human understanding, demonstrative syllogisms are the keystones of scientific knowledge. They must have starting points and such starting points are principles. By their nature of unprovable premises, however, principles cannot be demonstrated and this peculiar circumstance is what blocked the sceptics on their grounds, since they noted that it is impossible to trace back an infinite number of inferential steps, so it is impossible to deduce something certain from any chosen starting point if not posing it as just a mere assumption (*ἐξ ὑποθέσεως, Post. An*. I 3, 72b 7-15). But, according to Aristotle, *there are* some non-demonstrable propositions, i.e. first principles, that can avoid infinite regress *exactly because they are indemonstrable* and ground the scientific knowledge that eventually sprouts from them in the form of necessary inferences, for "either principles are demonstrable as principles of principles and the demonstration will go on forever (*εἰς ἄπειρον*), or the first starting points are [as they need to be] undemonstrable definitory items" (*Post. An.* II, 3, 90b 25-27). A demonstration that goes on forever cannot be taken as a proper demonstration because it indefinitely postpones its proof to an indeterminate future moment, if ever, so that along its way towards the expected conclusion we cannot even be sure whether such countless steps of inference actually constitute an epistemic progression or are just some sort of deductive but untargeted roaming. Also, the principle of non-contradiction, as a meta-principle, fits this schema because it cannot be deductively demonstrated, "for it is impossible that there should be demonstration of absolutely everything; <if it were so> there would be an infinite regress (*εἰς ἄπειρον γὰρ ἂν βαδίζοι*)", so that there would still be no demonstration (*Metaph.* Γ, 4, 1006a 6-10). Notoriously, according to Aristotle, the one who tries to refute the principle of non-contradiction will end up assuming it as mandatory for communication purposes, whatever reasonable will eventually be said. The dialectic that leads to the proof of the principle of non-contradiction leaves no further room for regressing to any antecedent or more foundational principle.

b) On the necessity of avoiding infinite regress *in good-oriented desires, evaluation process for deliberation, chrematistics and household self-sufficiency*

For Aristotle, the necessity of avoiding infinite regress also covers the sphere of human behaviours, so that strategical arguments that take potential infinite regresses into account can also be found in his *Ethics* and *Politics*.

[6] The classical example of Aristotle's denial of infinite regress is his idea of the highest good as something that is desired for itself. In fact, he

claimed that if we suppose that anything is good only if we desire it for the sake of something else, that is then good—this would yield to a regress that is never-ending and basically pointless. We should rather suppose that there must be at least one good thing that is not desired for the sake of something else that is good and that will be what has to be considered the highest good, i.e. some kind of fulfilment that is desired for the sake of itself. If, on the contrary, subjects of action always decided to make their choices for the sake of something else that is not oriented by what is good by itself, their aspiration would turn out to be perpetually unsatisfied and therefore idle and vain (*Nic. Eth.* I, 1094a 19-22). This argument is not meant to prove that all desires actually tend towards one last and superior aim, but to prove that all chains of desires tend to *some* final aim, whatever it may be, and that desiring only things that are good for the sake of something else makes no practical sense.

[7] This line of reasoning should be followed not only for the identification of our general target of desire, but also for the identification of contingent means of action. In other words, the evaluation process that leads to an operative choice cannot be infinite too. If the amount of evaluation that one has to do before taking any action consisted in considering every possible variable—an infinite number of variables—according to which things can go wrong (*Eud. Eth.* II, 1226b 1), then the requested evaluation would be at the very least impractical, because it would no longer be useful as it would (never) be brought to an end. That kind of pondering might go on forever and the evaluation would stop the initiative resulting in inaction, which is the opposite of what was expected at the beginning of the elaboration. According to Aristotle, the deliberative process must rather come to an end. The end is achieved when a choice is finally made by finding the most apt means to reach a previously identified aim, just as requested by the practical, practicable, and contingent purpose of the deliberative process itself (*Nic. Eth.* III, 1113a 2).

[8] Anti-infinitism should also be pursued in economics. As is known, there are two types of "chrematistics" or the art of enrichment and wealth-getting for Aristotle: household oriented or business oriented chrematistics. The former has simply the balanced management of the household as its unique scope, while the latter unlimitedly tends to the acquisition of wealth just for the sake of it, because apparently it has "no limit of the end" (*οὐκ ἔστι τοῦ τέλους πέρας*). Aristotle argues that even if this latter tendency is excessive and execrable because resources are limited and added financial value is not real value, sometimes it happens that the accountants of otherwise self-sufficient household management also devote their budgetary policy to unnecessary business. If that happens, it is likely that an

insatiable process of never-ending greed will arise, "They ought to increase their money without limit (*εἰς ἄπειρον*) or to consolidate it at any rate /.../ as their desires are unlimited, they also desire that the productive means of gratifying them should be without limit" (*εἰς ἄπειρον οὖν ἐκείνης τῆς ἐπιθυμίας οὔσης, καὶ τῶν ποιητικῶν ἀπείρων ἐπιθυμοῦσιν*. *–Pol.* I, 9, 1257b 25 - 1258a 14). But actual infinity is also impossible in this respect.

[9] Thus, regarding household management, Aristotle seems to be in favour of some moderate form of autarky oriented by an ideal of self-sufficiency. However, in consideration of the social and civic nature of men, such a self-sufficiency should never consist in total and absolute isolation of a single individual. On the other hand, he says, it should also not consist in some broad and too-inclusive network of partners, for "there must be some limit found also here: if our self-sufficiency is extended to ancestors and descendants and loved ones' loved ones, an infinite series will result "(*εἰς ἄπειρον*), so that the resulting interdependence will no longer be proper self-sufficiency. He specifies that "the self-sufficiency we posit as being what independence makes life desirable and lacking in nothing –and we think happiness is like this" (*Nic. Eth.* I, 1097b 7-21).

c) On the necessity of avoiding infinite regress in *sense-perception and soul-partitioning*

Some psychological analyses presented in Aristotle's *De Anima* also require an unbreachable bottom-line against infinite regress. Let us take the case of sense-perception first.

[10] Aristotle defends the thesis according to which all five senses of the human body guarantee self-awareness of what they perceive *and* at the same time self-awareness of perceiving what is being perceived. If that were not the case, then there should be an additional sense that guarantees awareness of the undergoing perception. For instance, if there were a sense of sight as separate from a sense of self-awareness of undergoing seeing, then there would be a problematic overlapping of receptors, because the former would perceive, e.g. colour, while the latter would perceive colour plus the perception of that colour that has already been perceived. As Polansky rightly wrote,

> /…/ were a further sense necessary for the awareness of vision, an infinite regress of senses begins—perceiving vision needs its own sense and perceiving that we perceive vision necessitates another, and so on—unless some sense is aware of itself, but then why allow the regress to begin at all rather than have the original sense perceiving itself. (Polansky 2007, 382)

Perhaps, if we follow such regression, sooner or later we might find a sense that can perform the two operations together, e.g. seeing and being aware of seeing, but the more plausible and economic explanation is rather that it is the sense in question, e.g. the sense of sight which provides vision, that realizes both operations by itself right from the beginning of the perception process. If a conscious sense needed to be activated by an infinite number of levels of self-awareness, then the perceiver would never be actually aware of his own perceiving—and for Aristotle this is clearly not the case.

[11] On a more general level, the organic soul as the unifying principle of the living body also cannot proliferate in levels of redundant complexity. Even if the soul can be subdivided in functional parts, it will nonetheless be possible to individuate some unifying structure that holds together the whole on a superior level than those exerted by particular psycho-physiological functions. Such a bond is not carried out by the body, since it is rather its immanent soul that holds together its functional unity. Therefore, if there was something more cohesive and more capable of holding together the soul than the soul itself, then such a presumed over-organic, meta-psychic principle should be called "soul" even more properly than the soul itself. "But again [Aristotle remarks] we shall have to repeat for it the question: Is it one or multipartite? If it is one, why not at once admit that *the soul* is one? If it has parts, once more the question must be put: What holds *its* parts together, and so *ad infinitum*?" (*De An.* I 5, 411b 9-13). The soul's division into parts cannot mean that those parts are spatially separable. Such a division is just potential and organized by a functionalist distinction that is always arranged on the basis of the *unity* of the living body. A limitless soul, infinitely fractioned in itself, cannot act as the principle of organic unity for living beings.

d) On the necessity of avoiding infinite regress in *ontological kinds or the criticism of the "Third Man" Argument*

A classical case where the insurgence of an infinite regress sets the basis for a *reductio ad absurdum* is Aristotle's criticism of the so-called "Third Man" argument.

[12] There, the object of Aristotle's critique is the presumed separation of the common predicate (intended as a Platonic Form, e.g. the Man as such) from the multiplicity of particular items (e.g. many men) from which it seems to over-emerge. If there really was something existing *as separate* from its referents, Aristotle argues, such a Form would lose its alleged ontological bond with the multiplicity that originates it. Thus we obtain a

kind that, in so far as it is distinct and separated from the reality it arises from, it is no longer instantiated by its sensible individuals. In other words, the Form of Man becomes abstracted from actual men, whose perceptual reality made the emergence of such a collective Form possible. So the question that arises is: What does still keep the Form of Man bound to its sensible, individual men? At that point, one would need a third concept, namely a more general and different set—a "Third Man"—, that can hold the Form of Man with men in flesh and bones together. But again, how to, in a comprehensive way, account for such a separate and over-ordered Form that participates both in particular men and the Form of Man? Another even more general idea would be required, and so then a fifth idea, a sixth idea and so on. Instead of a natural picture of reality, we end up with empty copies of copies, like the image that we obtain by placing a mirror in front of another mirror. Apparently, infinite regress was intrinsic in the idea of a separate Form as such, so this explicative model should be rejected from the very start. For Aristotle, it is impossible to separate the universal kind from its underlying individuals, so for him there is no chance of starting any regress at all (*De Ideis*, Fr. 4 Ross; see *Metaph.* A, 9 990b 16-23).

[13] In Aristotle's terms, it would be wrong if one presumed that essence and those beings that have such an essence were two different and separate things. As an absurdity, if essence were separated from that of which it is the essence of, then there would be two different things that, in so far as they are not the same, would require two different essences: "if they were different, the process would go on to infinity; for we should have the essence of one (*τὸ... τί ἦν εἶναι τοῦ ἑνὸς*), and the one (*τὸ ἕν*), so that in their case also the same infinite regress would be found. Clearly, then, each primary and self-subsistent thing is one and the same as its essence" (*Metaph.* Z, 6, 1032a 2-6). As a particular being cannot have more than one essence, Aristotle concludes, so it cannot be something different and separate from its own essence. [14] Another version of the Third Man argument is later provided by Aristotle in order to avoid infinite regress from a semantic point of view too (*Soph. El.* 22, 178b 37 - 179a 11). There he says that every time when a qualified individual (such as for instance a "Coriscus being literate") is taken to be diverse and existing separately from its respective singular individual (*"a certain this"* like "Coriscus", independent from his being literate or not), an ambiguity arises and some misunderstanding follows. Here the emerging infinite regress is both a clue and the evidence of this fallacy. For, if we suppose to individualize both "Coriscus" and "Coriscus being literate", we will be in need not only of hypothesizing a separate and superior type of unity for the two, but also for as many separate instances as the various qualifications of Coriscus and their combinations are in number.

Those separate instances turn out to be exponential *ad infinitum* as soon as one tries to find a more general and comprehensive over-set in this regard. As a particular case of missed type/token distinction, such syntactic proliferation is met for Aristotle whenever a predicative compound is taken individually and considered to be separately existent from its constituent meaning. Therefore, it is advisable not to distinguish and divide "Coriscus" from "Coriscus being literate". If we do so in our speech, we must be disposed as well to divide him, conceptually speaking, in an infinite number of instances.

e) On the necessity of avoiding infinite regress in parentage-filiation lines, in natural coming-to-be, and in generation processes in genera.

[15] According to Aristotle, all generation processes abide by the following law: "if a thing is being generated, there must be (1) *something from which* (*ἐξ οὗ*) *it comes to be* and (2) *something by which* (*ὑφ'οὗ*) *it is generated*, and this [something from/something by] cannot be anything that is infinitely [set back to other] (*εἰς ἄπειρον*)" (*Metaph.* Γ 5, 1010a 20-22). So, in order to realize some kind of generation it is at the least necessary to have both a passive condition realized, like some raw and elementary matter ("something from which generation comes"), and a non-passive condition realized, like some kind of agent that exerts upon such a pre-existing substrate triggering, the process of generation ("something by which generation comes"). In Aristotle's terms, these prerequisites are the material cause and the efficient cause respectively. For him, any accomplished and self-subsistent generation is proof that both the material cause and the efficient cause cannot be *infinitely* set back to other antecedent principles in a perpetual process of undergoing change. In particular, this argument was addressed in opposition to defenders of Heraclitean thought, according to which reality remains in a constant state of perpetual change and impermanence where every being sooner or later decays into non-being. Against them, Aristotle first noted that the extension of reality is not limited only to the corruptible and sensible world, where all the phenomenical changes take place (among mathematical entities, for instance, any change of that sort occurs), and, second, that the process of generation and corruption always has both a beginning and an end as its limits, so that change itself cannot be perpetual flow as Heracliteans maintained.

[16] In *Physics*, Aristotle further strengthened his point, assuming a transgenerational perspective:

> /…/ the process will go on to infinity if there is to be change of change and generation of generation (*μεταβολῆς μεταβολὴ καὶ γενέσεως γένεσις*); thus if

> one of a series of changes is to be a change of change, the preceding change must also be so: e.g. if simple becoming was ever in process of becoming, then that which was becoming was also in process of becoming, so that we should not yet have arrived at what was in process of simple becoming but only at what was already in process of becoming in process of becoming. And this again was sometime in process of becoming, so that it is not yet in process of becoming in process of becoming. And—Aristotle continues—since in an infinite series there is no first term, here there will be no first stage and therefore no following stage either. On this hypothesis, then, nothing can become or be moved or change [if that were the case]. (*Phys.* V, 2, 225b 34 - 226a 6; see *Metaph.* K 12, 1068a 33 - b 4)

The anti-infinitist argument, originally addressed against Heracliteans, is here extended to Parmenides and his advocates. The process of change and generation *de facto* requires some *actual* originators, who, like now-existing parents, cannot be considered as the perpetual offspring of previous ancestors in a never-ending line of ascendency. For Aristotle, if that were the case, then nothing could actually be generated. Instead, every actual and witnessed generation (*τὸ γιγνόμενον*) must be taken as the evidence that such an explanatory model is untenable. Since something has been generated (and we see it), it *had to have been generated* from the beginning to the end of its generation process. Clearly, this holds for natural generation as well as for every other type of physical change.

[17] It is rather the case that sensible entities shall also have some stable principles, i.e. their specific forms that are embedded in their matter and transmitted in the process of generation. Perishable things just becoming from perishable things is a domino-effect scenario, where a start cannot be found, so one might doubt the domino-effect as such. As Aristotle wrote,

> if principles [of generation] are perishable, it is clear that they too necessarily depend on certain principles (for all things perish into that out of which they are constituted), so that it follows that there will be other principles prior to those principles, but that is impossible, both if [the regress] stops and if it goes on ad infinitum. (Metaph. B, 4, 1000 b 24-28; see K, 2, 1060a 34-36)

It is then impossible that principles of generation are perishable for three reasons: 1. if they were perishable, they would have to have anteceding principles, and so the latter would not be proper principles; 2. if it were the case that it is possible that the anteceding regress may have come to a halt at some point, then such an arrest would no longer constitute a perishable principle as assumed; 3. if it were the case that the causal chain of perishable principles would not have an actual moment of beginning, then that would

be impossible because the "domino-effect" is ongoing in front of our eyes and it is proved wrong by the existence of living generations.

[18] The way in which the natural world works seems to support Aristotle's view, as he exemplifies at the beginning of his treatise *On the Generation of Animals*, where he imagines what would happen to zoological offspring if all types of crossbreeds were possible (*De Gen. An.* I 1, 714b 7-16). If all types of crossbreeds actually met, then the reproduction chances of the animal kingdom would reach a point of irreversible entropy. For example, we know that if a horse couples and mates with a donkey, then the offspring will be an equine hybrid, namely a hinny or a mule, depending on the parents' assortment. As hybrids, both the hinny and the mule are sterile in the majority of cases, so they are not able to reproduce themselves. Aristotle was concerned with what would happen if that were not the case and generative variations were exponentially possible: if hybrids were ever able to crossbreed themselves resulting in different offspring, that would be a problem for the survival of their progenitor species and for the persistence of their own derivative species as such, because in that case an excessive variety of progeny could have been generated, up to a point where the zoological differences among the descendants would be so huge that inter-coupling among the resulting new-borns would be no longer fecund. For this reason, according to Aristotle "nature flies from the infinite" (*φύσις φεύγει τὸ ἄπειρον*), because an infinite number of potential matches might lead to an uncontrolled proliferation of hybrids of hybrids, that in their turn might end up endangering the original reproductive species up to a point of mass self-extinction because of induced infertility. The peremptory sentence that concludes this argument—"infinite is imperfect while nature always seeks an end"—can be also translated as "infinite has no aim while nature always has one" (*τὸ μὲν γὰρ ἄπειρον ἀτελές, ἡ δὲ φύσις ἀεὶ ζητεῖ τέλος*), meaning that species' self-preservation is the aim that animals seem to have, while infinitistic crossbreeding seems to hinder this purpose.

f) On the necessity of avoiding infinite regress at the microscopic level, in the process of material dissolution and in the number of mono-elementary interactions

[19] According to Aristotle, minimal size should not be the only criterion of individuation when one looks for the minimal corporeal elements that are at the bedrock of sensible reality. For him, it is not just the physical dimension, however microscopic, that defines and distinguishes what a basic element is from what is not. For, "those whose ground of distinction is size will have to recognize an element prior to the element, a

regress which continues infinitely, if it is true [as it is] that every body is divisible and that which has the smallest parts is the element" (*De Caelo* III, 5, 304b 7-9). The process towards infinite minuteness can be never-ending, whatever the smaller element under analysis is. *Let us take the example of "fire", Aristotle says. Fire is commonly considered the thinnest and most rarefied element (λεπτότατον), like a material corpuscle but with minimal consistency. Consequently, it is arguable that the constituent components of such an evanescent body, if existent, are the most rarefied ones (λεπτομερέστατον). Indeed,* Heracliteans believed that *fire was precisely the material element that was composed of the most subtle, impalpable and weightless constituents, especially if in comparison to other material elements and their presumed ulterior constituents. Following this line of reasoning,* they believed that the world in its entirety was ultimately made of *fire because fire was the element to which everything amounted to if reduced to its basic constituents.* However, this is not how it works, Aristotle remarked, because we better define basic material elements not by their alleged minimal size (which as physical bodies are infinitely divisible in potency), but rather according to the fundamental interactions that those presumed building blocks of reality (i.e. elements) can undergo in virtue of their non-recessive potentialities.

[20] Not to mention that, if fire were the only original element of reality and all the other basic elements would come to existence after it, then some impossible consequences would follow as well. If everything came from fire, then everything would comply with the inner ascending tendency of that originating element. This means that everything would tend above to find its collocation in the natural place of the fiery element, i.e. in the superior concavity of the elementary sphere. If that were the case, then all the subsequent material transformations would occur in the same place and in reference to just one common conforming element. What would happen in that case is that in change after change, the environment would be impoverished and depleted of all variety of material interactions, up to a point where everything existing would be simple fire and in a condition of perpetual rest, or better, perpetual tendency to the higher reachable location. That would happen because fire by itself can neither have all the physical qualities, nor guarantee all the oppositions (hot/cold and dry/wet) that are fundamental for the reciprocal and cyclical transformation of matter in the sublunary world. Aristotle foresees that "an infinite number of contrarieties would attach to the single element" so that "at the end elements are not further transformed into one another"; but for him "the process must come to a stop, and cannot continue *ad infinitum* in a straight line in either direction" (*De Gen. et Corr.* II, 5, 332b 6-16). By way of example, it would

be like a painter who wants to change the shade of his primary colours over and over using a palette which contains only the colour red: the outcome will be that all the paintings will turn out just plain red. Clearly, the same discourse can be applied to the other external element, i.e. earth, whether considered to be the originating one at the expense of the others, as well as for the other elements whose natural tendency is to place themselves between the extremes, i.e. air and water. Reality would turn out to be mono-elementary, Aristotle says, while phenomenical experience seems to give us evidence that the ongoing changes constantly occurring in the sensible world are provided by the natural interactions of non-reciprocally-reducible elements. Those elements, on the other hand, cannot be infinite in number, otherwise an illimited quantity of interactions would take place so that sensible reality would be physically indeterminate and inconclusive.

[21] Another related issue is how to conciliate the perdurance of material elements (that seem to be always present and inexhaustible in their becoming) with the fact that such material elements undergo constant dissolution because of their constitutive corruptibility. Apparently, a way to account for this incompatibility would be to say that dissolution itself is never-ending, as if such a process were constituted by continuous and endless steps of decay. But Aristotle estimated that if the process of decay were never-ending, that would require an infinite time to reach the point of complete break up, i.e. when constitutive elements return to be the next available ones for new material aggregations. But again, if the intervening dissolution interval were infinite, then new aggregations could never take place. Moreover, since parts of wholes have different times of dissolution, different times of infinite dissolution would start in unsynchronised moments, so that the same would happen for their infinitely postponed compositions. The envisaged prospect is paradoxical at least: "It will follow that there are two infinite times which are mutually exclusive, the time occupied by the composition, which is infinite, being preceded by the period of dissolution [that is infinite as well]. There are thus two mutually exclusive infinites, which is impossible" (*De Caelo* III 6, 304b 27- 305a 1).

g) On the necessity of avoiding infinite regress in presuming the creation of matter and form at the time of production or generation of individual compounds, as well as providing definitions by defining their parts

[22] For those who, like Aristotle, reject *creatio ex nihilo,* when something comes to be then something must already be given. This requirement holds true both for generation of natural beings and for production of artificial objects. According to Aristotelian hylomorphism in particular, the formal

structure (*εἶδος*), as well as the material substrate (*ὕλη*), must be both available *before* the actual generation or production of individual substances takes place, otherwise an infinite regress will emerge. An example may serve the purpose of explanation. Let us make reference to the Aristotelian example of the production of a bronze sphere, whose matter is bronze while its form is the sphere. When one shapes bronze in spherical fashion, what usually happens is that a selected quantity of metal is melted and poured into a predisposed mould that can withstand the cooling and hardening of metal in the chosen form. Both the bronze (or matter) and the spherical mould (or form) must be given before the forging process starts. In the acts of generation or production, neither is the form originated nor is the matter created for the first time (see *Phys.* I 9), but rather a formal cause as well as a material cause must *pre-exist* to be merged together in the resulting compound (*σύνολον*), where they start their co-existence in that singularity. If, on the contrary, we presume that form is created along the process of individual production or generation, Aristotle says, then we will fall into an infinite regress of antecedent pre-production or pre-generation requirements.

> For if we make the form [within the process of production], we should make it out of something else; for this was assumed [since form is always enmattered]. For instance: we make a bronze sphere; and that in the sense that out of this, which is bronze, we make this other, which is a sphere. If, then, we create the [form of the] sphere itself, clearly we must make it in the same way] i.e. out of something that is shaped by such form], and the processes of making will regress to infinity. (Metaph. Z 8, 1033a 34- b5)

The creation of a bronze sphere would be an impossible task to accomplish if these were the conditions, because it would require as preliminaries the antecedent creation of the form of a sphere, which would in turn require some anterior bronze to be enmattered as needed, which would in turn require a prior anterior form of sphericity to be realized, and so on. Thus, within the process of individual production or generation, not everything is being created, even less so by a series of retrograde creations. We should look at the bronze sphere once it is done—when nothing is happening anymore: "neither the matter nor the form comes to be—i.e. the last matter and form. /.../ The process, then, will go on to infinity, if not only the bronze comes to be round but also the round or the bronze comes to be; therefore there must be a stop at some point" (*Metaph.* Λ 3, 1069b 35 - 1070a 4). Here the reference to the "last" matter and form (*τὰ ἔσχατα*) should be explained as reference to *the most proximate* matter and form, namely those that are concomitantly assumed by the compound once it is made and completed at the end of the actual process of its production or generation.

The finished bronze sphere is the empirical proof that the making of individual items cannot be subject to endless regressive involution that has been presumed in the making of its structural and substantial components.

[23] For Aristotle, becoming as such needs some irrevocable grounding, because for him there cannot be infinite substance or body, no infinitely extended space, no actually infinite magnitude (see *Phys.* III 5, 204b 24 ff.). Some grounding is also required along the definitory research of universal unities. For the parts of which definitions are made of (e.g. "rational" and "animal" as *definiens* of "man") should be self-explanatory and in any case should not have to be defined within the definition to which they are attached to, otherwise second-order definitions of *definientes* would arise, which would in turn be specified by other definitions of definitions and so on. As we already considered, definitions cannot be constituted by an infinite number of definitory terms (extensively and now intensively intended), since that would be useless and unpractical at the very least. In the same way, definitions of hylomorphic compounds do not have to account for *all* the parts that constitute the said compound, even less so all the material components (*Metaph.* Z, 10, 1035b 5-14). So, even if we know, for example, that a circle is composed of two semi-circles, we should avoid making reference to the definition of "semicircle" when we define what a circle is. It is rather the other way around: semicircle is in fact defined in reference to the entire whole that it is part of, i.e. as half of it. Only by maintaining an adequate level of generality will definitions not fall into the regress of over-specification, because if circles were defined in terms of semicircles, then semicircles could be defined in terms of quarter-circles of which they are composed, and so on *ad infinitum*. If that were the case, providing the basic definition of "circle" would turn out to be a Sisyphean effort for any prodigal geometer. Thus, for Aristotle, essential definitions are composed of simple terms that are not definable further. Furthermore, from an etymological point of view, definition (*ὁρισμός*) is what sets a clear-cut boundary between what something is and what it is not by ruling out what is not pertinent and holding together what it is about: be that the boundary is permeable or limitless, then proper definitions would lose their exclusive peculiar function, making no more sense for the scientific purposes for which they were intended. The ampler they are, the more redundant and loose-fitting they will be. If it is impossible to provide a close-fitting definition on the specific level, then one may forever be looking for a wider-fitting one with no result (see *Metaph.* α, 2, 994b 16-20).

h) On the necessity of avoiding infinite regress within the property of "being-in-a-place", at the start of a physical series of moved movers and in the introduction of undue celestial spheres

[24] Avoiding infinite regress can be an easy exit strategy from apparent paradoxes or *insolubilia*. In that sense, Aristotle quoted and challenged the so-called Zeno's "aporia of place", an argument meant to show how the property of "being-in-a-place" could be intrinsically problematic. As is known, Zeno's spatial puzzle was intentionally formulated to raise a provocative paradox, where the alleged infinite regress played a fundamental role. In fact, it starts by saying that everything that exists is somewhere, but also that place is something that exists. Therefore, it turns out that place is also somewhere, so that in every identical location there are two existing things: the place and the thing placed in that space. But where are they? In just one and equal place. Consequently, each place and placed thing will have in their turn a spatial place where they are placed, and so on *ad infinitum*. We may imagine that Zeno even tricked his interlocutors, forcing them to admit that it would then have been impossible to say where the place of a thing really is, up to the point that one might even wonder whether such a thing was actually placed somewhere or rather *nowhere*. If place had its existence "inside" placed things (like in a Russian nesting doll model of space), an infinite series of inherences would open up, so that, for instance, spatial translations would turn out to be impossible. In fact, in Zeno's scenario, nothing could move out, say, from its bottomless spatial pit placed right within itself. However, even if it is true, Aristotle conceded, that "being-somewhere" can mean "being-in-something", this does not mean that "being-in-something" reflexively means "being-within-a-place" as Zeno intended. "Place" for Aristotle is rather something that extends all across the surface covered by the dimensions of the placed object *without being part of the object itself*. Therefore, the relation between content and container can always be explained in a non-recessive, but dispositional way. For Aristotle, a given body which contains its content, e.g. an amphora containing wine, acts as a *limit* of that content, without being part of its content. "So we escape the infinite regress", Aristotle concludes, and the paradox of unplaced places instantly vanishes (*Phys.* IV 3, 210b 27).

[25] Paradoxes excluded, some of the most famous infinite regress arguments provided by Aristotle undoubtedly remain those aimed at showing that causal successions without origin are physically impossible. Movement, for instance, is a telling example.

> Let us take the case [Aristotle says] in which a thing is in motion and is moved by something that is itself in motion, and that again is moved by

> something else that is in motion, and that by something else, and so on continually; then the series cannot go on to infinity (μὴ βαδίζειν εἰς ἄπειρον), but there must be some first mover. (ἀνάγκη εἶναί τι τὸ πρῶτον κινοῦν. – Phys. VII 1, 242a 50-53)

In fact, according to one of the fundamental principles of his mechanics, *everything in movement must be moved by something*, but motion transmission cannot take place without a start. At a global as well as at a local level, the causal chain of the efficient causes that pushes the world in motion cannot proceed as it actually proceeds without the input of some prime mover. Clearly, such a prime mover has to not be moved in its turn (no longer being *the* starter though), because a regress will occur and the overall movement will never have its actual beginning. The "domino-effect", so to speak, that transmits motion across the universe must have a triggering inciter somewhere. As he wrote,

> /…/ since in an infinite series there is no first term (πρῶτον), it is impossible that there should be an infinite series of movers each of which is itself moved by something else /…/ [For] there must be something that imparts motion not by something else but by itself (αὑτῷ), or else there will be an infinite series (εἰς ἄπειρον). If, then, something moved incites movement, <even so> the series must stop somewhere (ἀνάγκη στῆναι) and not be infinite. (Phys. VIII 5, 256a 17-19; 27-29)

Like a mechanism whose gearwheels need an ever-turning power-engine to be set in motion, to work and to transmit their derivative motion to subsequent gearwheels, Aristotle's cosmos needs a first mover (which moves only by means of itself) in order to function as it constantly does.

[26] Such a cosmological meta-fulcrum does not exclude the repetition of motion transmission but rather allows it, prompting the eternal circular turning of celestial spheres. Aristotle, as Eudoxus before him, wondered about how many celestial spheres one should presume are in the heavens. For, one might think that it makes no difference how many celestial spheres there are once their prime mover has been identified as their agent. However, since celestial spheres were supposed to carry around their respective heavenly bodies, their number must be equal to the number of observed bodies arranged by common movement (e.g. one sphere for the fixed stars, one for Saturn, one for Jupiter, etc.). If there were more celestial spheres than those that are needed for carrying around the existent heavenly bodies, they would move aimlessly without any purpose, just for the sake of themselves. But, again, Aristotle's guiding-principle says that "nature does nothing in vain" and this also holds economically. Regardless of the number of observed celestial common motions, introducing redundant

celestial spheres is for Aristotle tantamount to introducing redundant translations of translations. At this point, however, nothing prevents one from thinking that the actual number of celestial spheres is *countless*. But even if one imagines an infinite number of celestial spheres turning around in the heavens, this would amount to having an infinite number of transmissions of transitions that would never be brought to an end, so that inferior bodies, as for instance the moon, would never be seen in transit, but would be still. Therefore, in this case Aristotle also concludes that "if a movement is to be for the sake of a movement, this latter also will have to be for the sake of something else; so that since there cannot be an infinite regress, the end (*τέλος*) of every movement will be one of the divine bodies which move through the heaven" (*Metaph.* Λ 8, 1074a 29-31). That is to say that those planetary and starry regular transitions are the end-points of all the celestial spheres one might imagine, so it is convenient and appropriate to stick to their equivalent number without adding further levels of unobserved concentricity.

Conclusion: One Last Infinite Regress Argument Against Actual Infinity

To conclude, another Aristotelian argument can be briefly recalled that sums up most of the infinite regresses examined so far. This last, comprehensive passage can be considered as a sort of "etiological" *manifesto* against actual infinite regress through infinite regress arguments (*Metaph.* α 2 is the case in point here).

As we have seen, for Aristotle a series of causes cannot be infinite in any way. Now it is specified that this holds both for the number of causal successions and for the number of the kinds in which particular causes are arranged. For, not only the efficient cause, but also the material, the formal, and the final ones cannot express an infinite quantity of inner cause/caused steps along their course. All of them must not have an infinite number of intermediate passages, otherwise they would always be inconclusive. As said, inconclusiveness is proved impossible enough by phenomenical, sensible reality. The four Aristotelian causes must rather have some *limit* at the extreme of their consequential chain. Such a limit shall not only be in the ascending line, i.e. by having a *first* causing-cause, but *also in the descending line*, i.e. with a *last* caused term. For, Aristotelian causal connections always spread out between an initial *terminus post quem* and a final *terminus ante quem* that might be temporary but fully-present at some point. If the head of the causal chain did not exist, the causal chain as such would not exist. And if there were infinite intermediate terms, none of them

would actually be *the* first, so that again the causal chain, without a start, would not exist as such. On the other hand, there cannot be a lack of one last term or actualized effect, since every cause is cause *of* something and there cannot be anything like an "uncausing cause".

Temporal succession, however, is not the only parameter that is at play in cause/caused relations, as in, e.g. the day that comes after the dawn. For, according to Aristotle, causal generation can take place in two different ways, both of which reject infinite regress, given that one singular cause cannot have an infinite number of offspring.

[27] One way is irreversible and can indeed have intermediate phases, although those phases are within a *continuum* and finite in number independently of how one wants to divide them. For example, "being man" as caused by "being child" is an irreversible process (it is impossible to turn it back) and has intermediate phases (e.g. puberty, adolescence, and manhood). Sooner or later, at some time during the process, "being child" will have caused "being man". That condition is an end-point in the *continuum* towards maturity from which there is no turning back. As far as one wants to split the phases of growth, the actualized man still constitutes the limit and the aim (*τέλος*) against the pretended infinite involution and towards which "being a child" tends and finishes its causal extension.

The existence of a limit and an aim is the same reason why the final cause cannot be infinitely undetermined. As we have considered in the case of the impossibility of infinite procrastination, the absence of a term of desire determines the collapse of any purpose in the decision-making process. And on a more general level, the outcome would be that neither the reason-why nor the good would exist as such. For Aristotle, however, that is in patent contradiction with the common behaviour of men, who usually tend to pursue what they believe is good on the grounds of their knowledge and desiderative capacities. Should the process be infinite and the effort not worth the trouble, living beings would not even give it a try. But, again, that is not the case of how it works in practice, and infinite regress here seems to be repudiated by custom.

[28] The other way of causal derivation has an end-point as well, but has no defined intermediate phases and is reversible, even if it implies the dissolution of the cause. For example, the element of "air" as caused by "water". Here infinite regress is avoided by the circularity of the process, since water can always return to its form after air will be transmuted into fire, fire into earth, and earth back into water. The "never-endingness" of the cycle of elements does not imply depletion, but rather its being environmentally self-sustaining. At the same time, "water" is the end-point of the causal chain that brought it from "air", and that also holds for each of

the other elements at their stage of stability. The high term of this process is first matter, which is the permanent substrate of all sensible stuff and is always steady. While the elements, insofar as they are the proximate causes of coming-to-be as such, can undergo change, corruption, and cyclical regeneration without interrupting the ongoing transmutative process. For Aristotle, they even *dissolve* one into another, because that precisely allows change and continuity without stagnation or regression. Such a dissolution for change is what prevents the physics of elements from a potentially infinite dispersion along this cyclical process of renewal.

[29] In regard to the formal cause, we have already considered what would happen if definitions were limitless or wider than needed. It is worth stressing, however, that for Aristotle an infinite chain of formal causes, as is the case of the Third Man argument, would impede the possibility of scientific knowledge as such. Concept analysis and concept transmission would be impossible to attain because open-ended definitions would be unmanageable for anyone wanting to establish something irrefutable or indisputable on their grounds. Aristotle here faces a captious objection, since one may argue that, on the contrary, geometers firmly know what a line is because they know infinite divisibility as its peculiar properties. Such objection is captious because it introduces infinity within the process of knowing, so that if it were received, then actual knowing might be had of any infinite subject-matter. However, Aristotle famously replied that even if the geometer knows that the line can be infinitely divided, no-one actually knows, nor can actually list or perform, the infinite divisions of the line. Moreover, one (like Zeno) would be wrong in presuming that a *geometrical* line can be infinitely divisible as a *sensible* being in motion can be. Epistemic definitions are given of non-recessive objects only, and, even if they are mathematical objects, without the need to list all their potential parts, so that there is even no need of knowing infinity in all its actual instances (that would be impossible by the way) to have finite knowledge of infinity. Aristotle's incisive conclusion here sounds like a sort of principle against actual infinite regress: "nothing infinite can exist, and if it could, at least being infinite is not infinite (*οὐκ ἄπειρόν γ'ἐστὶ τὸ ἀπείρῳ εἶναι*)" (*Ibid.*, α, 2, 994b 26-27. See *Phys.* III, 6-8). In other words, "the infinite can only be seen through the finite, the latter being the only form of existence of the former" (Drozdek 2008, 101).

[30] For Aristotle, the inquiry about reality as oriented by the four causal queries, the formal being just one of them, covers the understanding of the world in full and exhaustively. In the framework of his scientific endeavour, no other kinds of queries are needed and no less than those four (*Metaph.* A). Accordingly, there is never the risk of getting lost in the never-ending

indeterminateness of actual infinity nor in some unfounded scepticism. All questions can be reduced to these and all the essential explanations will be best addressed by them. Here "Aristotle's razor" is particularly sharp and cutting. It is even pointless for him to speculate about a potentially infinite number of causal typologies. After all, causal queries, however fundamental, are human artefacts. Therefore, the "horizontal" number of kinds of causes cannot be infinite because man, being with the quality of the subject of their knowledge, is a structurally finite being with cognitive limitations. Man conceived of these questions through his finite mind. He could not be capable either of producing nor grasping an infinite list of reasons, since "it is impossible that one can traverse an infinite series in a finite time" (*Ibid.* α, 2, 994b 30-31. See *Phys.* VI, 7, 230a 31-36). Therefore, the approach that is called "infinitism" in contemporary epistemology (namely the pretension that knowledge may be justified by an infinite net of reasons) would be complete non-sense for Aristotle. What is highly consolatory for him is rather the certainty that, sooner or later, an end *can* be found both at the beginning and at the end of each causal chain. Rational animals have just to look for those ends as prescribed by their specific nature.

Bibliography

Bonitz, Hermann. 1870. *Index Aristotelicum*. Berlin: apud G. Reimerum, s.v. "*ἄπειρον/*progressus in infinitum".

Bostock, David. 1973. "Aristotle, Zeno and the Potential Infinite". *Proceedings of the Aristotelian Society*, Vol. 73: 37-51.

Bostock, David. 2012. "Aristotle's Philosophy of Mathematics". In *The Oxford Handbook of Aristotle*, edited by. Ch. Shields, 465-491. Oxford: OUP, Oxford University Press. In particular, § 4: "Aristotle on Infinity".

Bowin, John. 2007. "Aristotelian Infinity". *Oxford Studies in Ancient Philosophy*, Vol. 32: 233-250.

Cambiano, Giuseppe. 1989. "Aristotele e i disagi dell'infinito". In *L'infinito dei Greci e dei Romani*, 27-47. Genova: Dipartimento di archeologia, filologia classica e loro tradizioni.

Cameron, R., 2018. "Infinite Regress Arguments". In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Accesed December 14, 2020. https://plato.stanford.edu/ archives/fall2018/entries/infinite-regress/.

Charlton, William. 1991. "Aristotle's Potential Infinites". In *Aristotle's Physics: A Collection of Essays*, edited by L. Judson, 129-149. Oxford: Clarendon Press.

Coope, Ursula. 2012. "Aristotle on the Infinite". In *The Oxford Handbook of Aristotle*, edited by Ch. Shields, 267-286, Oxford: Oxford University Press.

Corclius, Klaus. 2017. "Aristotle and Thought Experiments". In *The Routledge Companion to Thought Experiments*, edited by M.T. Stuart, Y. Fehige, and J.R. Brown, 37-76. Abingdon: Routledge.

Côte Antoine,.1990. "Aristote admet-il un infini en acte et en puissance en *Physique* III, 4-8?". *Revue Philosophique de Louvain*, Vol. 88: 487-503.

Dowden, Bradley. 2013. "Infinite". *Internet Encyclopedia of Philosophy*. Accessed December 14, 2020. https://www.iep.utm.edu/infinite/.

Drozdek Adam. 2008. *In the Beginning was the Apeiron: Infinity in Greek Philosophy*. Stuttgart: Steiner.

Edel, Abraham.1934. *Aristotle's Theory of the Infinite*. New York: [s.n.].

Falk, Ruma. 1994. "Infinity: A Cognitive Challenge". *Theory and Psycology*, Vol 4: 35-60.

Furley, David J. 1969. "Aristotle and Atomists on Infinity". In *Naturphilosophie bei Aristoteles und Theophrast IV*, edited by I. Düring, 85-96. Heidelberg: Lothar Stiehm.

Giardina, Giovanna R. 2012. *Fisica del Movimento e teoria dell'infinito. Analisi critica di Aristotele*, Phys. III. Sankt Augustin: Academia Verlag.

Gratton, Claude. 1999. "Viciousness and Actual Infinity in Aristotle's Infinite Regress Arguments". In *Proceedings of the 4th International Conference of the International Society for the Study of Argumentation*, edited by F. H. van Eemeren, R. Grootendorst, J. A. Blair, and Ch. A. Willard, Amsterdam: SIC SAT. Available at: http://rozenbergquarterly.com/issa-proceedings-1998-viciousness-and-actual-infinity-in-aristotles-infinite-regress-arguments/.

Gratton, Claude. 2010. *Infinite Regress Arguments*. Dordrecht: Springer.

Hintikka, Jaakko. 1966. "Aristotelian Infinity". In *Philosophical Review*, Vol. 75, No. 2: 197-218. Reprinted in *Time and Necessity. Studies in Aristotle's Theory of Modality*, 114-134. Oxford: Clarendon Press.

Hussey, Edward. 1993. *Aristotle's Physics Books III and IV* (revised ed.). Oxford: Clarendon Press.

Ierodiakonou, Katerina. 2011. "Remarks on the History of Ancient Thought Experiment". In *Thought Experiments in Metodological and Historical Context*, edited by K. Ierodiakonou and S. Roux, 37-49. Leiden-Boston: Brill.

Johnstone, Henry.W. Jr. 1996. "The Rejection of Infinite Postponement as a Philosophical Argument". *Journal of Speculative Philosophy*, Vol. 10: 92-104.

Jori, Alberto. 2010. *Das Unendliche. Eine philosophische Untersuchung*. Norderstedt: Books on Demand. In particular pp. 11-44.

Kouremenos, Theokritos. 1995. *Aristotle on Mathematical Infinity*. Stuttgart: Steiner Verlag.

Kucharski, P. 1954. "L'idée d'infini en Grèce". *Synthèse*, Vol. 34: 5-19.

Lear, Jonathan, 1979. "Aristotelian Infinity". *Proceedings of the Aristotelian Society*, Vol. 80: 187-210.

Mariani, M., 1994. "Il 'Terzo Uomo' nelle Confutazioni Sofistiche". In *Atti del Congresso Logica e Filosofia della Scienza: problemi e prospettive*, edited by C. Celluci, M. C. Di Maio and G. Roncaglia, 3-13. Pisa: ETS.

Mondolfo, Rodolfo. 1934. [1968, 2012]. *L'infinito nel pensiero dei Greci*. Le Monnier, Florence.

Polansky, Ronald. 2007. *Aristotle's De Anima: A Critical Commentary*. Cambridge: Cambridge University Press.

Post, John F. 2010. "Infinite Regress Argument". In *A Companion to Epistemology* (2nd ed.), edited by J. Dancy, E. Sosa and M. Steup, 447-450. Oxford: Wiley-Blackwell.

Prevosti Monclús, Antonio. 1985. *La Teoría del Infinito de Aristóteles*. Barcelona: Promociones Publicaciones Universitarias.
Rescher, Nicholas. 2010. *Infinite Regress. The Theory and History of Varieties of Change*. New Brunswick-London: Transaction Publishers.
Stewart, Ian. 2017. *Infinity. A Very Short Introduction*. Oxford: Oxford University Press.
Sweeney, L. 1960. "L'infini quantitatif chez Aristote". *Revue Philosophique de Louvain*, Vol. 58, No. 60: 505-528.
van Steenberghen, Fernard. 1985. "Aristote et l'infini". In *Aristotelica: Mélanges offerts à Marcel de Corte*, edited by C. Rutten and A. Motte, 337-350. Bruxelles-Liège: Presses Universitaires de Liege.
von Fritz, Kurt. 1969. "Das *apeiron* bei Aristoteles". In *Naturphilosophie bei Aristoteles und Theophrast IV*, edited by I. Düring, 65-84. Heidelberg: Lothar Stiehm.
Virno, Paolo. 2010. *E così via, all'infinito. Logica e antropologia*, Torino: Bollati Boringhieri, 2010 partial translation by L. Chiesa in Virno 2011.
Virno, Paolo. 2011. "The Antropological Meaning of Infinite Regression". *Angelaki. Journal of the Theoretical Humanities*, Vo.16: 63-76.
Waldegg, Guillermina. 1993. "El infinito en la obra aristotelica". *Educación Matemática*, Vol. 5, No. 3: 20-38.
Weidemann, Hermann. 2017. "Potentiality and Actuality of the Infinite: A Misunderstood Passage in Aristotle's *Metaphysics* Θ.6, 1048b 14-17". *Phronesis*, Vol. 62, No. 2: 210-225.
Weinert, Friedel. 2016. *The Demons of Science. What They Can and Cannot Tell Us About Our World*. Cham: Springer. In particular, Chapter 2: "Thought Experiment in Ancient Greece", 9-16.
White, Michael J. 2009. "Aristotle on the Infinite, Space, and Time". In *A Companion to Aristotle*, edited by G. Anagnostopoulos, 260-276. Oxford: Wiley-Blackwell.
Wieland, Jan Willem. 2013. "Infinite Regress Arguments". *Acta Analytica*, Vol. 28: 95-109.
Wieland, Jan Willem. 2014. *Infinite Regress Arguments*. Springer: Cham.
Wieland, Wolfgang. 1962. *Die aristotelische Physik: untersuchungen uber die Grundlegung der Naturwissenschaft und die sprachlichen Bedingungen der Prinzipienforschung bei Aristoteles*. Gottingen: Vandenhoeck und Ruprecht.
Wolf, Sebastian. 1983. *Das potentiell Unendliche: die aristotelische Konzeption und ihre modernen Derivate*. Frankfurt am Main - Bern: P. Lang.
Zellini, Paolo. 1980. *Breve storia dell'infinito*. Milano: Adelphi.