

## University of Groningen

### Examining Life

van Mazijk, Corijn

*Published in:*  
Second Thoughts

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2021

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

van Mazijk, C. (2021). Examining Life: An Introduction to Phenomenology and Existentialism. In M. Sie, & B. Engelen (Eds.), *Second Thoughts: First Introductions to philosophy* Open Press TiU.

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Tilburg University

### Second Thoughts

Sie, Maureen; Engelen, Bart

DOI:

[10.26116/secondthoughts-sie-engelen-openpresstiu-2021-1](https://doi.org/10.26116/secondthoughts-sie-engelen-openpresstiu-2021-1)

*Publication date:*  
2021

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Sie, M., & Engelen, B. (Eds.) (2021). *Second Thoughts: First Introductions to Philosophy*. (1 ed.) Open Press TiU. <https://doi.org/10.26116/secondthoughts-sie-engelen-openpresstiu-2021-1>

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Second Thoughts

First Introductions to Philosophy

*An open educational and open-ended handbook*

Maureen Sie & Bart Engelen (eds.)

# SECOND THOUGHTS

First Introductions to Philosophy

*(An open educational and open-ended handbook)*

Maureen Sie & Bart Engelen (eds.)

DOI: 10.26116/secondthoughts-sie-engelen-openpresstiu-2021-1

ISBN: 9789462406636 (Interactive PDF)

**Published by:** Open Press TiU

**Contact details:** [info@openpresstiu.edu](mailto:info@openpresstiu.edu)

<https://www.openpresstiu.org/>

**Layout Design by:** Wolf Publishers, Claudia Tofan & Kaftwerk

**Artwork/Portraits:** S. Lloyd Trumpstein

**Cover Photo Credit:** Philadelphia Museum of Art: The Louise and Walter Arensberg Collection, 1950, 1950-134-59

© Artists Rights Society (ARS), New York/ADAGP, Paris/Succession Marcel Duchamp

Open Press TiU is the academic Open Access publishing house for Tilburg University and beyond. As part of the Open Science Action Plan of Tilburg University, Open Press TiU aims to accelerate Open Access in scholarly book publishing.

The **TEXT** of this book has been made available Open Access under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 license.



# CONTENTS

## **Second Thoughts**

---

*Introducing the Introductions*  
Maureen Sie & Bart Engelen

## **Thinking about Societies**

### **Power, Justice and Identity**

---

*An Introduction to Political Philosophy*  
Constanze Binder

### **Why Ethical Reflection Matters**

---

*An Introduction to Moral Philosophy*  
Maureen Sie and Bart Engelen

### **Imagined and Contested Communities**

---

*An Introduction to Philosophy of Culture*  
Thijs Lijster

### **Religious Beliefs: Can they Be Legitimate?**

---

*An Introduction to Philosophy of Religion*  
Herman Philipse

## **Thinking about Humans**

### **Cartesian Conundrums**

---

*An Introduction to Philosophy of Mind*  
Monica Meijsing

### **Plastic Mortals and 21st Century Healthcare**

---

*An Introduction to Philosophical Anthropology*  
Jenny Slatman

### **Appreciating Beauty and Art**

---

*An Introduction to Aesthetics*  
Bart Vandenabeele

### **Examining Life**

---

*An Introduction to Phenomenology and Existentialism*  
Corijn van Mazijk

## Thinking about Thinking

### **The Quest for Truth**

---

*An Introduction to Epistemology*

Tim De Mey

### **The Bounds of Reason**

---

*An Introduction to Philosophy of Science*

Chris Buskes

### **The Study of Informational Processes**

---

*An Introduction to Logic*

Sonja Smets and Fernando R. Velázquez Quesada

### **'You can't argue with that!'**

---

*An Introduction to Argumentation*

Catarina Dutilh Novaes

## Thinking about Reality

### **This is the Question: To Be or Not to Be?**

---

*An Introduction to Ontology*

René van Woudenberg

### **The Highest Principles? The Death and Rise of 'First Philosophy'**

---

*An Introduction to Metaphysics*

Eddo Evink

### **Language and Reality**

---

*An Introduction to Philosophy of Language*

Menno Lievers

### **Artwork Acknowledgements**

## **SECOND THOUGHTS**

### ***Introducing the Introductions***

Maureen Sie & Bart Engelen

### **INTRODUCING PHILOSOPHY**

**P**hilosophy, in its many forms and guises, has a broader relevance than many people realize. In societies where populism reigns, political philosophy can help you appreciate the meaning, value and importance of democracy. When conspiracy thinking becomes widespread, epistemologists and philosophers of science can help you understand what it means for something to be true and count as knowledge. One might even argue that the quicker the pace of our lives, our technologies, our news and our decision-making processes, the bigger the need for the kind of careful, systematic and critical thinking that characterizes philosophy.

Because that is exactly what good philosophy is: thinking slowly, reflecting critically, realizing that things may not be as obvious as they might seem at first hand, trying to figure out how things hang together from a broader perspective. Philosophy, we believe, is a verb. When you read good philosophy, this encourages you to actually *do* philosophy.

Doing philosophy might be confusing at first, introducing questions where you initially thought you had answers. However, an advantage of this journey is that you will end up on firmer ground. You will know better than before where you stand, why you believe the things you believe. But you may also come to know what you did not know after all. This last realization, strange as it might sound, might be valuable as well: it makes room for intellectual humility. Carefully

weighing the arguments for different positions and views with respect to a certain issue and thoroughly examining the assumptions behind these positions and views help you grasp their complexity, how they relate to each other and why people may continue to disagree with each other. Once you grasp that things are complex, that there are many things we do not fully understand and that people may have very different approaches to the same issue, you can start to build up from what you do know, what you do grasp, and what you do agree on.

Thinking critically and questioning things thoroughly does not mean that you will end up with the view that anything goes, that there is no such thing as the truth or that mere opinion rules. Relativism, as the view just articulated is often called, is only one of many philosophical positions about truth and knowledge and whether we can attain those. This is, in fact, one of the topics discussed in the chapters of this handbook. Other questions addressed are: What exactly does justice require? What is the right thing to do? Is there something that unifies a culture? What is human nature? What do we mean when we say that something is beautiful? What are the limits of science? How exactly does language influence our thinking? Why is there anything at all? These questions matter, even if most of us live our lives without knowing the answers to them or fully grasping those answers.

Philosophy, in a sense, is about having second thoughts. When we have second thoughts, we allow room for doubt about things we thought we knew. As Socrates and René Descartes already stressed, doubt can be considered the eminently philosophical method. When we do philosophy, we doubt and wonder about what amazes and surprises us and we are led by our curiosity to see where our thinking may lead us. When we listen to scientists, journalists, politicians and, granted, also some philosophers, they all too often seem to be very sure of themselves, having all figured out how the world works, how we should act and what the truth is. This handbook, in contrast, aims to stimulate doubt and to find out for yourself whether they are right. It all starts with having an open mind and being willing to question the preconceptions of yourself and others. Both in our education and in our societies at large, these philosophical attitudes and skills are valuable things to have.

This handbook is illustrated with well-known portraits of philosophers, adapted and beautified by Tilburg-based artist S. Lloyd Trumpstein (a.k.a. Stephan de Weert). The cover of the book is a painting by Marcel Duchamp that dates back to 1912 and is officially called '*Nude Descending a Staircase, No. 2*'. When painting it, Duchamp himself claimed to be aiming for "a static representation of movement, a static composition of indications of various positions taken by a form in movement." While in some sense static as well, this handbook aims to do the same thing: capture something that is inevitably and crucially in movement, namely our thinking about ourselves, our societies, our worlds and our thinking itself.



## INTRODUCING THE HANDBOOK

This handbook is an open educational and open-ended resource for whomever is interested in philosophical thinking. Each of the chapters is open in the sense of freely available and accessible to everyone. You may be a student who wants to get some background on a specific philosophical sub-discipline. You may be a teacher who wants to assign introductory reading for students. You may be a layperson interested in reading an overview of philosophical thinking, written in a clear and accessible way. Each of you: feel free to browse, download, print and use the collection as you see fit. We believe that open access is the future and that academic philosophy as presented in this volume is of potential worth to many of you out there.

In this open-ended handbook you find two kinds of chapters. First, there are chapters that provide a broad introduction into a specific philosophical sub-discipline, such as political philosophy, epistemology or metaphysics. As this collection covers most of the sub-disciplines currently taught at Western philosophy departments, you can legitimately claim that you have been introduced to Western 'philosophy' as a whole, understood rather canonically, after having read the entire handbook. Second, there are chapters that introduce slightly more specific topics or philosophical approaches. You will always be able to know the focus of each chapter by looking at its subtitle.

The open-ended nature of this handbook, means that new chapters will be added in the future. We hope that philosophy will change and grow with time to become more diverse and inclusive and that this handbook will do so as well. We think of philosophy and its evolution as an organic process, as a tree that branches out in many different directions, adding new directions as it goes along. In this handbook, we organize the wide variety of topics that philosophers discuss into four main branches, which represent important subject areas that philosophers have covered.

First, there is 'thinking about societies', which includes chapters that cover philosophical approaches to matters of obvious societal relevance. How should we organize our societies? How should we treat others? What exactly are cultures and what role do they play in a globalized world? This branch covers philosophical discussions, theories and views on what binds and divides us as societies and communities.

Second, there is 'thinking about humans', which includes chapters that zoom in on people, the members that make up those societies. Is there something like human nature and what does that look like? How do human minds and bodies relate to each other? Are we free or not? This branch covers what one could broadly call 'philosophical anthropology': philosophical discussions, theories and views on what it means to be human.

Third, there is ‘thinking about thinking’, which include chapters that focus on the ways in which humans can relate to the outside world. How can we come to know things about that world? What is truth exactly? What are the values and limits of scientific understanding? How do we reason and argue and how do we do so properly? This branch covers philosophical discussions, theories and views on how humans come to believe things about themselves and the worlds they live in.

Fourth, there is ‘thinking about reality’, which includes chapters that investigate those worlds in more direct ways. Do things have an essence? What do we mean when we say that some things exist and others do not? How can language help us access the reality out there? This branch covers philosophical discussions, theories and views on the world we, as humans, find ourselves in.

If you like what is on offer in this handbook, you can let us know on the website <https://www.openpresstiu.org/> and register for updates, for example when new chapters are added.

Consider each chapter as a first and stand-alone introduction to the exciting and thought-provoking world of a specific branch of philosophy. The same will be true of future chapters. Like the chapters already included, these future chapters will be accessible for readers without any specific prior knowledge. All you need is curiosity, an open mind and a willingness to think twice.

## **ACKNOWLEDGEMENTS**

**L**ike philosophy, this handbook has its own history. Its roots lie with Martin van Hees and Lodi Nauta, who planned for a collection of introductory chapters that would together make up for a philosophical handbook introducing readers to academic philosophy. While their plan did not, for a number of reasons, materialize, they gathered a lot of the authors who now fill these pages of this first edition, as we could call it, of this handbook. We thank them for their work and the collegiality when the project was handed over to us.

We also want to thank the authors of this first edition for their contributions and their patience in this process. When we took over as editors, their chapters were read by yet another pair of eyes and all authors gracefully collaborated with us to see things through. While each has their unique voice, the collection as a whole, we hope, is coherent in terms of style, accessibility and setup of each chapter.

We also want to thank everyone involved in reading the chapters. That includes, again, Martin van Hees and Lodi Nauta, but also the many colleagues and native speakers who you will see mentioned in the first footnote of each chapter. We thank Ronja Rönnback for assisting us in the early stages of our project.

Last but not least, we want to thank Daan Rutten for offering us the opportunity to publish this in what we believe should be the future of academic science and education: fully open access. If you now have this freely available and interactive PDF at your disposal, this is in part due to Daan's efforts to promote open access. As editor of Open Press TiU, he has also been extremely helpful in the final steps of this process: editing, typesetting and designing the manuscript that you are now looking at.

---

# THINKING ABOUT SOCIETIES

---

Constanze Binder

**Power, Justice  
and Identity**

# POWER, JUSTICE AND IDENTITY

## *An Introduction to Political Philosophy*

Constanze Binder<sup>1</sup>  
[binder@esphil.eur.nl](mailto:binder@esphil.eur.nl)

### 1. INTRODUCTION

Imagine you could re-design society, say on the regional, national or even global level, without yet knowing which role you would have in it yourself. What would you propose? Which values and rules, if any, should govern society? Who should rule? What would be the form of government? How can we (if at all) justify the monopoly of power of a state and what government would you choose? How would you distribute economic resources? What (if anything) do we owe future generations and how should this affect our current action, for instance, in the light of climate change or pandemics like corona?

Political philosophy is an area of philosophy and political theory that aims to answer these questions. It draws on an age-old tradition of thinking that focuses on fundamental questions about the organization of human collaboration. An important feature is its normative nature. Political philosophy focuses not so much on how things are but rather on how they ought to be.

At the same time, philosophers are also “children of their time”. That is, their thinking and the questions they raise are influenced by the developments of the

---

1. I am very grateful to Martin van Hees and Lodi Nauta for their detailed editorial comments and very useful suggestions to further improve this chapter. Furthermore, I would like to thank Wiep van Bunge and Han van Ruler for their support in tracing historical sources employed in this chapter.

world surrounding them, for instance the civil war that inspired Thomas Hobbes to write his *Leviathan*, the intellectual climate shaped by Rousseau and other Enlightenment thinkers that led to the French Revolution, the Holocaust that shaped Hannah Arendt's life, or the global inequality and poverty that Martha Nussbaum and Amartya Sen address in their work. This is one of the many fascinating sides of political philosophy: it stands on the shoulders of giants, drawing on the insights of some of the greatest thinkers of (Western) history, but it also allows one to extend and use these insights to shed new light on the burning issues of one's own time.

Of course, philosophers are also human beings: they have, like all of us, some remarkable blind spots in their thinking, at least so we think from our own vantage point. For Aristotle, for example, the exclusion of slaves and women from the status of democratic citizenship was seemingly so self-evident that it did not seem to require much justification. Similarly, Mary Wollstonecraft, a pioneering 18th century thinker inspiring the movement for women rights to vote, was convinced that the right to vote should be restricted to women of status, leaving them enough time for the required study to engage in politics.

In this light, an engagement with political philosophy is a fascinating and challenging endeavour but vulnerable to the blind spots that characterise our own time and thinking. Developments are going fast and seem often to be unpredictable. Ten years ago, hardly anybody could have predicted the UK leaving the EU or the election of a business man and TV star as president of the United States of America. "Fake news" is new on the political agenda, while other topics have become prominent again, such as the increase in inequality of economic and political power of the more affluent members in our societies.

Why are such tendencies puzzling? Do they refer to developments, as some claim, that negatively affect or even threaten the cornerstones of democracy? Wouldn't a government of wise leaders or experts, such as economists or health care specialists, be better suited to manoeuvre society through difficult times of financial crisis or pandemics? Can a society with economic inequality as we now experience it be stable? Can it be just? Can our democracy deal with these developments or do we need to rethink the way we organise our societies to account for the fundamental pre-conditions for a democracy to work?

We start in section 2 with questions about the foundations of our societies and the justification of the state. We then turn, in section 3, to different forms of governments and the reasons to value democracy. Section 4 discusses the notion of justice and how some of the most influential political philosophers of the last century envisioned a just society. Questions about how polarisation in liberal societies may have given rise to populism and the return of nationalist

sentiments are the subject of section 5. We conclude in section 6 with an outlook of possible directions that political philosophy might take in the future.

## 2. JUSTIFYING POWER

One of the core questions with which political philosophers are concerned is: can one justify the existence of the state and, if so, which types of government are to be preferred? Our world is organised in nation states. Of course, we can read up in history books how the current nation states came into being and how the boundaries on our maps were drawn. However, this does not yet answer the question of how we can justify the existence of states and the transfer of power to those who rule them. What gives the (moral) grounds for the monopoly of power that governments usually possess? What can possibly justify the power to include some and exclude others from admission to a nation state?

Philosophers have often thought about the justification of the state by means of a thought experiment in which one imagines a “state of nature”, that is, a world in which the state does not exist and in which its inhabitants are thus unconstrained by legal rules or external authority. How would such a world look like, how would it be to live in it? Our answers to these questions will affect our assessment of the state. If the state of nature is a situation in which our lives would be miserable, then we may prefer to live in a world where the state does exist.

Hobbes (1651) famously argued so. He thought the state of nature would be a situation of a war of all against all: without the protection and security offered by the state, rivalry over scarce goods would lead to conflict and ultimately to war. As result, the individuals in the state of nature will agree that a state is needed and they will, so Hobbes argues, transfer their power to it. According to Hobbes, the state is justified because it is needed to maintain peace and civil order and because citizens (hypothetically) consent to it.

Is this argument plausible? Hobbes’s pessimism about the prospects of living peacefully in a state-less society was based on a particular view about human beings. Locke (1689) uses the same thought experiment but is more optimistic about the possibility of human cooperation in the state of nature. Yet, he also concludes that the individuals will decide to form a state and defer some of their powers to it. Such a state possesses a monopoly of force and the protection that it gives to citizens is taken to be superior to the messiness and inconvenience of a state of nature in which individuals themselves try to settle their conflicts.



However, neither Hobbes's nor Locke's argument is undisputed. Some anarchists have argued that a peaceful state of nature is possible and that it is to be preferred to the possible abuses of power committed by states. Other defendants of anarchy see the role of the state merely as temporary. The state is necessary until humans have perfected their conduct and are ready for anarchy, that is, the state is a necessary preparation for living without it.

Obviously, the argument from a state of nature is a thought experiment - we have never actually set up or consented to a government in this way. Nevertheless, the argument is useful in our understanding of the possible justificatory grounds of nation states. If, contrary to the anarchists, one does agree with Hobbes and Locke and takes the thought experiment to give sufficient justification of the state, then the next question is: who should govern us and how should they do so?

### 3. WHO SHOULD HAVE THE POWER?

The question of who should govern or rule us can also be put in this way: Which form of government should we adopt in our society? Should the power be completely in the hands of one person (dictatorship) or a small group (oligarchy) or should it be "a government of the people, by the people, for the people" (democracy)? I shall limit myself here to a discussion of the justification of democracy.

When inquiring into the meaning and prerequisites of democracy, one of the first difficulties one encounters is that the term is often used in different ways. Sometimes it is simply a form of window dressing, such as with the former German Democratic Republic (or East Germany) or the Democratic Republic of Congo. But when we consider what we do take to be genuine democracies, we encounter quite some variety, making it difficult to come up with a definitive definition. To discuss the many models and versions of democracy that have been put forward over the centuries falls of course outside the scope of this chapter. Instead, we shall look into some of the main reasons as to why democracy is valued. Getting a clearer view on these reasons will provide us with a guide to assess different democratic systems in different contexts and thus possibly also clarifies the demarcation between democracies and other forms of government.

It is helpful to start with an argument that has been raised *against* democracy: an examination of this argument sheds light on why we take democracy to be valuable. According to Plato (367 BC), a government of experts - philosopher-kings as Plato calls them - trained in a detailed and extensive way would be best suited to rule. Only the wise man knows what is best, and the wise man should thus rule. This view presupposes that there is such a thing as a 'best' decision

in politics. But what makes a political decision good or bad? One answer is to say that a good decision traces “the common good”. A possible objection to this view, i.e. that a wise ruler would be best, is to say that under certain conditions democracies would actually do a “better job” in arriving at the common good than a government of a group of experts, be it philosophers or others. Furthermore, people may disagree about what the common good is or be uncertain about their interest. In the face of such uncertainty, it may well be that the “wisdom of the crowds”, and thus democracy, is a better way of arriving at the truth than following a singular person or group.

The utilitarians (further discussed in the chapter on ‘moral philosophy’) Jeremy Bentham and John Stuart Mill took the common good to consist of the happiness (or “utility”) of all people. Though this definition is far from uncontroversial, there are aspects of life that are usually considered to be an undisputed part of the common good, such as people’s livelihood and health, which are better realized in democracies. The economist Amartya Sen has famously argued that famines, one of the greatest threats to people’s health, have never occurred in a functioning democracy. Similarly, some argue that democracies are more peaceful and less likely to engage in wars than other forms of government such as autocratic rule.

Whether democracies do indeed perform better in these respects, can in the end only be established empirically. But this reveals a potential problem. It suggests that democracy is to be grounded on what philosophers often call ‘instrumental reasons’: we value democracy only as an instrument to achieve other goals such as people’s happiness or stable economic growth.

Are there no reasons why we value democracy for its own sake? Is there an *intrinsic value* to democracy, that is, a value independent of its effects? A conversation which this author had with a Tunisian activist of the Arab Spring movement brings this question about the intrinsic value of democracy to light: this activist made clear that while supporting the protests in the beginning, he was now convinced that Tunisians were worse off than before, giving both the economic situation and overall security issues as reasons for a decrease of wellbeing and happiness. He thought therefore that it would have been better to stick with the previous regime rather than to move to democracy. What would you answer him? Suppose you would want to defend the merits of democracy beyond its possible positive impact on our happiness, how would you argue for it?

One prominent intrinsic value attributed to democracy is connected with the importance we ascribe to human freedom and autonomy. More specifically, if we understand freedom in the way of the French philosopher Jean-Jacques Rousseau, namely as “the obedience to a law we give to ourselves”, then irrespective of its outcomes, democracy is to be cherished because it best

realizes this freedom. We find it important to have such control ('autonomy') irrespective of whatever further benefits we derive from it.

One can object that autonomy need not be secured in a democracy. We still have to obey laws with which we may disagree. Moreover, it can be the case that the same minorities are structurally outvoted by the majority. What to make of these possibilities? Do they undermine the defence of democracy in terms of its intrinsic value? Not necessarily. Some will argue that the members of minority groups have an equal possibility to influence the law-giving process: to be outvoted need not mean a lack of influence. Others argue that a democracy is not only characterized by majority voting but also by the allocation of civil and political rights. These rights should ensure the absence of abuses of power and allow citizens to pursue their own ideas about how to live their lives as long as it will not interfere with the rights of others to do so as well.

This leads to another aspect of the intrinsic value often attributed to democracy, namely equality. Irrespective of the outcome of democratic decision-making procedures, all people can participate in them equally. Be it the "one person, one vote" principle in representative democracy or the equal chance to convince one's co-citizens in public debates, adult citizens have the same possibility to participate in the democratic procedure. Of course, one might question what such equality requires in order to fully realise it in practice. If better educated people or those with higher incomes have easier access to the political process, will we still say that equality is realized? Or what about possible unequal chances to cast an informed vote if information is not sufficiently available in the language of a country's minority group? The answers given to these and similar questions might differ, but they concern what is required to realise democratic equality and underscore its importance.

Even though the exact nature of the intrinsic value that we attach to democracy is thus open for discussion, its importance shows that a defence of democracy need not only refer to its presumed beneficial effects in terms of economic prosperity, peace building or overall happiness. Even if democratic regimes do not improve and in fact would even worsen the economic outlook or the political stability in a country, there are powerful arguments to defend democratic institutions. This, however, does raise the question how to respond to situations in which security or health of the people can be served more effectively by putting democracy, even if only temporarily, on hold. Does a terroristic threat justify a state of emergency that limits civil liberties, as happened in France after the Paris attacks in 2015 for over two years? Similarly, in case of a fundamental health threat as in the Covid-19 pandemic, can states legitimately suspend freedoms, such as the freedom of movement or assembly, and to bypass or suspend regular democratic procedures?

Remaining vigilant and aware that especially in times of fear, be it terrorism or a virus, we may all have a tendency “to rally around the flag” and accept crucial cuts in our civil and democratic liberties that we otherwise would not accept. As Arendt persuasively reminds us, the conditions and dynamics away from democracy are never inevitable. Being aware of the mechanisms that can undermine democracy makes it possible to address and overcome such threats.

In section 5, we will discuss a recent threat to democracy: the declining relevance of facts and truth in politics. But before we go into this, we first analyze what many also take to be another threat:—the economic inequality between the world’s richest 1% and the rest (as prominently identified by French economist Thomas Piketty). Inequalities in economic power and wealth can undermine political equality and democratic processes. This raises the question when the distribution of social and economic goods can be said to be just and when not. What are the fair terms of cooperation among members of society? In the next section, we will see how this question, which was perhaps the central theme of political philosophy in the last five decades, is addressed.

#### **4. JUSTICE AND A FAIR DISTRIBUTION OF THE FRUITS OF COOPERATION**

**T**hinking of society as a system of cooperation, as discussed in section 2, raises the question on how the economic and social benefits of a society’s joint endeavours should be distributed. What should be the basis or criteria of justice? How should we distribute our resources? How much inequality is still acceptable, if at all? Should everyone have the opportunity to pursue their idea of what they take to be the good life? If so, what is required for it? Indeed, how should we start reasoning about this question and how should we make sure that personal interests do not bias our judgements about justice?

The philosopher who has made a huge impact on our thinking about these questions is John Rawls, whose *A Theory of Justice* (1971) is one of the most influential philosophical works in the twentieth century. He asks how we can reason about justice and come to an agreement on which institutions characterize a just society. One problem that we often encounter in discussions about this question, is that ideas about justice might be biased by one’s social position, one’s upbringing, status or other characteristics such as gender or religion. How can we then come to an impartial account of justice, that is an account that allows us to say which societal arrangements and institutions are just, irrespective of our position in society (and the effect such arrangements have on us personally)?

Rawls solves this problem with a thought experiment. He postulates the so-called original position, that is, a hypothetical situation in which members of

society discuss which institutional arrangements would be just. The people are behind a “veil of ignorance”, that is, they do not know which position they will have in the future society that will be governed by the principles they are to agree on; neither do they know the class to which they will belong, their gender, religion, culture or other aspects of their life. Because of this lack of information, people in this original position will think and reason about the rules of society without yet being able to know how they will affect themselves. According to Rawls, this means they necessarily will be impartial when thinking about the principles of justice that should guide their society.

Rawls argues that the parties behind the veil of ignorance would agree on two main principles that institutions in a just society should satisfy:

1. Each person is to have an equal right to the most extensive total system of basic liberties compatible with a similar system of liberty for all.
2. Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged, and (b) attached to offices and positions open to all under the conditions of fair equality of opportunity.<sup>2</sup>

These two principles of justice are ordered lexically, meaning that one first has to safeguard the most extensive set of equal basic liberties for all. The next step is the realization of the second principle, which ensures that all have an equal opportunity and the means necessary to build, revise and pursue what they themselves take to be worthwhile striving for in life, safeguarded by the basic liberties. According to the Rawlsian principles of justice it can thus be just if people in some positions in society earn more, as long as those positions are open to everyone (equal opportunities) and as long as paying them more also benefits the least advantaged people in society. Say if higher earnings for some in society leads to an increase of overall wealth that is used to increase minimal wages.

Next to a lot of praise and admiration, Rawls’s theory of justice has also received quite some criticism. One criticism voiced by his Harvard colleague Robert Nozick is that the Rawlsian principle of distributive justice focuses on the wrong things. To see whether a distribution is just, says Nozick, we should examine how it came about, not whether it fits a certain equality ideal or some other pattern.

To modify a famous example employed by Nozick, consider a society (say societal state A) where resources are distributed according to whatever principle you subscribe to. Say you believe that equality of wealth is of overriding importance and let A be a situation in which this equality is completely realized. A famous football player, call him Ronaldo, has bargained with his team that

---

2. Rawls, John (1971), *A Theory of Justice*, Cambridge: Harvard University Press; pp. 302-3.

every visitor to a match in which he is to play next season, will pay an extra euro for him only. His team mates are happy with the arrangement because it allows them to play with one of the greatest football players and the management is also happy with it because of the success it promises to bring the team. People love to see Ronaldo play, and are enthusiastically paying the extra money to see him. When the season is over, the equal distribution of wealth is no longer existing - Ronaldo ends up much richer than other citizens. If the same scheme repeats for a number of years, inequality will have vastly risen (societal state B). Why, so Nozick asked, would state B be problematic if this arose by voluntary transactions only? To assess whether a distribution is just, we must, so Nozick argues, not look at its specific features but to the way it came about.

Nozick formulates a principle of justice: if a distribution is just, then any subsequent outcome that results from voluntary actions only also is just. In his *Anarchy, State and Utopia* (1974), Nozick developed this and other principles (e.g. one concerning the justice of the initial situation and another principle about how to deal with violations of justice) that do not rule out widespread inequality. In line with the second (and possibly even the first) principle of justice defended by Rawls, one would support redistributive policies if such redistribution improves the situation of the worst off. But Nozick would empathically oppose such a policy, arguing that such taxation would amount to "forced labour".

Another crucial point of criticism of the Rawlsian theory came from a different angle: disagreeing with Nozick and assuming that equality is important, Amartya Sen famously asked: what is it exactly that we find important to equalise for the sake of justice? Different philosophers have given different answers to this question: utilitarians would reply that everybody's utility needs to be weighted equally, while others focus on, among others, income and wealth. Sen highlights various problems with these answers. Some people's tastes are very expensive, and it is unfair to say that we have to make sure that they are just as happy as those who are easily satisfied. With respect to financial means, Sen points out that personal features of a person may greatly affect what people can actually do with it. A disabled person can, for instance, do fewer things with the same income as another person. Personal characteristics such as age, physical condition, gender, or the social and environmental aspects of one's surroundings, may affect how one can use one's economic means. Simply ensuring for everyone to have the same financial means may thus still create huge inequalities between people. It neglects human diversity and the corresponding differences in what one needs to pursue the ends that one values in life.

As a consequence, Sen introduced the notion of capabilities. Capabilities are the freedoms that people have reason to value and that are actually open to them. These freedoms are substantive in the sense that they refer to actual opportunities people have. Thus, the freedom to move around is something that

we all have reason to value but it may require different means for a disabled person. Together with Martha Nussbaum, Sen developed this approach into what is now known as the “capability approach”, and which forms a distinct way of thinking about equality and justice.

The question about the ‘currency of justice’ (happiness, income, wealth, and capabilities by no means exhausts the list), has led to a rich debate about distributive justice. Although the political implications of the views of Rawls, Nozick, Sen and Nussbaum and others on justice differ greatly, and range from the ‘left’ to the ‘right’ on the political spectre, the views are all seen as ‘liberal’ perspectives by taking individual rights, liberty and autonomy as core elements. The debate has also led to a renewed interest in non-liberal approaches to justice. Marxism, in particular, rejects the individualist perspective. I lack the space to discuss the Marxist alternative here and will instead turn to a more recent critique of liberalism.

## 5. IDENTITY, LIBERALISM AND POPULIST POLITICS

Rawls assumed that the people behind the veil of ignorance do not know the social affiliations and group identities they will have in the society they are designing. Some critics argued that this fails to account for the importance of people’s group membership, which - if ignored - leads to a problematic neglect of structures of domination or exploitation that people qua being members of particularly oppressed groups are subject to. According to these critics, Rawls implicitly projects the values of an intellectual male elite, e.g. such as rationality and individuality, onto the so-called neutral and impartial parties behind the “veil of ignorance”. By doing so he ignores the reality of disadvantaged groups such as women, black people or cultural minorities. Iris Marion Young (1990), for instance, argues that this neglect of taking people’s memberships in oppressed or disadvantaged groups into account leads to a continuation of the status quo; the Rawlsian theory re-enforces the position of the powerful.

To account for the realities of oppressed groups, the role and importance of recognition of different cultures has become more prominent in recent decades. How should a state with a core commitment to pluralism and state neutrality regarding the objectives that people pursue in life, deal with situations of oppression? Can it justify special requirements of particular groups? How can a framework identify and remedy oppressive structures that disadvantage people qua being members of a group? One answer given by Iris Marion Young herself, is to shift the focus from the alleged neutrality of Rawls towards a state that acknowledges and cherishes diversity in values and background.

Whereas the importance of group membership and the attention for possibly hidden patterns of oppression can be seen as a critique of the political 'left', the focus on people's identification with a group or culture (and the values and preferences coming with it) is also seen as major factor of a shift towards the populist politics of Trump and others. Both movements form instances of what is called 'identity politics': the appeal to the reality and demands of particular groups, whether these are the identities of oppressed (minority) groups (as in case of the left) or the mobilisation of the more conventional (nationalist) identities (as often, though not exclusively, exercised by the populist right). In both cases it is a shift that, according to Francis Fukuyama (2018), risks to undermine the old ideals of democratic societies by moving away from equality and freedom as core values, towards a more fragmented society in which politicians appeal to voter's identities.

More generally, what is common to nearly all populist politicians is that they usually claim to speak for or represent "the people", whose interests they claim to protect against elites or against the effects of migration or (economic) globalisation. A common interpretation of the unexpected rise in populist politics across the Western world is that an increasing number of people feel left behind by the conventional political parties. This analysis is shared by populists from the left and the right. The exact explanation may subsequently differ. Some refer to economic reasons and express the view that the process of globalization has led to gross inequalities. Others refer to presumed threats to one's traditional values and cultures, either by migration from "outside" or by new feminist or LGTB movements from "within". A return to nationalist rhetoric of putting one's own people first by restricting migration or trade belongs to the standard repertoire of such politicians as does the opposition to all kind of other societal changes.

The increasing polarization within society as a result of identity politics and populism, poses a serious threat to liberal democracies and, as some argue, to societies in general. One possible response,—according to Fukuyama, is to conceive of one's identity not in narrow terms, but in more inclusive and broader terms. One is not, say a white middle-aged female from the outskirts of Rotterdam, but a Dutch, European or, in the words of Desiderius Erasmus, a citizen of the world.<sup>3</sup> To populists, however, who argue for a return to nationalism or patriotism, such a broader 'cosmopolitan' perspective is just one of the many unwanted features of globalization.

It is not possible to predict to what extent the different views about the value of cultural diversity and inclusion will continue to dominate the future political

---

3. Erasmus, Desiderius (1522/2011), Letter to Huldrych Zwingli, letter 1314, in *De correspondentie van Desiderius Erasmus*. Deel 9: Brieven 1252-1355, translated by Jan Bedaux (2011), Rotterdam: Ad. Donker.



agenda. But it is obvious that it is one of the most important challenges to liberalism, both in its 'left' and 'right' versions, that political philosophy has helped shape the last decades.

## 6. CONCLUSION

Our societies are currently confronted with a number of developments such as climate change, pandemics, and migration that transcend the confines of the nation-state. As with populism, the philosophical and political questions which these developments raise, require political philosophers to move beyond the conventional horizons with which they were concerned until the end of the last century. Instead, they will have to broaden their perspectives from nation states to the world at large and from policies here and now towards inter-generational and environmental justice.

Global justice concerns issues of (distributive) justice on a global scale and the duties that states owe towards, for instance, refugees and migrants. Similarly, questions about inter-generational justice have become particularly prominent because of climate change and its effects. What do we owe our children, and our children's children and future generations more generally? Are we required to leave a share of the Earth's resources to them? Do our emission-intensive life styles create duties towards people in the global South, where emissions have started to increase only more recently? These are but some of the timely debates of today's political philosophers. Inspiration for answers can be found in the work of great thinkers of the past but the new developments and challenges that confront us now also require new answers and new approaches.

## RECOMMENDED FURTHER READING

- Kymlicka, W. (2001). *Contemporary political philosophy: An introduction*. Oxford University Press.
- Swift, A. (2019). *Political philosophy, a beginner's guide for students and politicians* (4th ed.). Polity Press.
- Wolff, J. (2016). *An introduction to political philosophy* (3rd ed.). Oxford University Press. (Original work published 1996)

## BIBLIOGRAPHY

- Arendt, H. (1951). *The origins of totalitarianism*. Schockens Books.
- Aristotle. (1984). *The Politics* (C. Lord, Trans.). University of Chicago Press.
- Erasmus, D. (2011). *De correspondentie van Desiderius Erasmus. 9: Brieven 1252 - 1355* (J. Bedaux, Trans.; 1. dr). Ad. Donker. (Original work published 1523)
- Fukuyama, F. (2018). *Identity, the demand for dignity and the politics of resentment*. Profile Books.
- Hobbes, T. (1651). *Leviathan: With selected variants from the Latin edition of 1668* (E. M. Curley, Trans.; 7th print). Hackett. (Original work published 1651)
- Locke, J. (1988). *Two treatises of government* (P. Laslett, Ed.). Cambridge University Press. (Original work published 1689)
- Mill, J. S. (2002). *Utilitarianism* (G. Sher, Ed.). Hackett Publishing Company. (Original work published 1861)
- Nozick, R. (1971). *Anarchy, state and utopia*. Harvard University Press.
- Nussbaum, M. (2011). *Creating capabilities, the human development approach*. Harvard University Press.
- Piketty, T. (2014). *Capital in the twenty-first century*, (A. Goldhammer, Trans.). Harvard University Press. (Original work published 2013)
- Plato. (2012). *Republic* (C. J. Rowe, Trans.). Penguin. (Original work published 367 B.C.E.)
- Rawls, J. (1999). *A theory of justice*. Harvard University Press. (Original work published 1971)
- Rousseau, J. (2003). *The social contract* (M. Cranston, Trans.). Penguin Random House. (Original work published 1762)
- Sen, A. (1999). *Development as Freedom*. Oxford University Press.
- Sen, A. K. (2001). Equality of what? In R. E. Goodin & P. Pettit (Eds.), *Contemporary Political Philosophy: An Anthology* (pp. 473–483). Blackwell. (Original work published 1980)
- Wollstonecraft, M. (2004). *A vindication of the rights of woman: With strictures on political and moral subjects*. Penguin Random House. (Original work published 1792)
- Young, I. M. (1990). *Justice and the politics of difference*. Princeton University Press.



HANNAH ARENDT

## HANNAH ARENDT

### THE ORIGINS OF TOTALITARIANISM

---

One of the merits of philosophy in general and political philosophy in particular is that it allows us to draw on insights of the past, to understand the present and change the future. Hannah Arendt's *The Origins of Totalitarianism* is an excellent example. Being of Jewish origin and a refugee fleeing Nazi Germany to Paris and New York thereafter, Arendt (1906- 1975) discusses in detail the historical conditions that enabled the horrors of the second world war. She masterfully describes the societal conditions on national and global levels for the rise of totalitarian rulers, as in the Nazi and Stalinist eras, to conclude that these provided fertile ground for and supported the rise of new forms of totalitarian regimes. The rise of such a regime however is not an inevitable outcome of history. Her message is that the end of dark times can mark a new beginning, in which political philosophy can sharpen one's critical view as a citizen and human being. It can lead to a renewed questioning of one's own thoughts, as well as of the *Zeitgeist* of the time, thereby contributing to the exploration of new beginnings and new directions. *The Origins of Totalitarianism*, as well as Arendt's later work, testifies in a brilliant way to the important role that critical reflection can have in countering negligence, apathy and thoughtlessness, thereby actively shaping and changing societal structures. Political philosophy is not the only tool to do so, but it is a very powerful one, not only within the universities but also, and perhaps more importantly, beyond the walls of these academic institutions.

---

Maureen Sie &  
Bart Engelen

**Why Ethical Reflection  
Matters**

# WHY ETHICAL REFLECTION MATTERS

## *An Introduction to Moral Philosophy*

Maureen Sie and Bart Engelen  
[m.m.s.k.sie@tilburguniversity.edu](mailto:m.m.s.k.sie@tilburguniversity.edu)  
[b.engelen@tilburguniversity.edu](mailto:b.engelen@tilburguniversity.edu)

### 1. INTRODUCTION

Imagine having a nice glass of wine on the terrace of your favorite restaurant, catching the last rays of summer sun. A long-time friend joins you because she recognized the restaurant in the picture you posted on social media. There you are, momentarily forgetting about your troubles and those of the world. But the next thing you know, your friend starts talking about ethics. What would you imagine her to start talking about? The topics that probably spring to mind are that the restaurant is serving too few vegetarian options or that it is bad (though in this case fortunate) that with all these social media around, there is so little privacy left. Or perhaps your friend feels guilty about spending money on a fancy dinner after reading about people starving in other parts of the world. Nothing spoils a leisurely moment more than talking about moral issues, or so it seems.

While vegetarianism, privacy and global poverty are indeed important topics in moral philosophy, ethics covers many other aspects of our lives as well. Many of our everyday decisions, even very mundane ones, are suffused with moral considerations; we constantly choose on the basis of things we value or care about, even though we might not always realize that is the case. A lot of the time, moral issues remain implicit, and we take values, such as friendship, for granted. When we do explicitly discuss these, we are often addressing a topic that somehow bothers us, however slightly. It bothers us because the norms

and values we tend to take for granted can come into conflict with one another (may I spend my money the way I please, or do I have obligations toward others?) or can leave us empty handed when deciding what to do.

Moral philosophy, or ethics, which we take as synonyms, is the systematic and critical reflection on what is good and bad, right and wrong, morally speaking. It is concerned with articulating and examining moral reasons that concern us all because they govern our everyday interactions, whether we are aware of them or not. Ethics is concerned with examining both the internal coherence of our moral reasons and their strengths and weaknesses in justifying how we lead our lives and run our societies.

In this sense, moral philosophy is perhaps closest to political philosophy, which not only aims to understand what politics is about (a descriptive enterprise), but also reflects on what it should be (a normative enterprise). In this last respect, political philosophers have recourse to fundamental values such as freedom, equality and justice.

In what follows, we introduce the three most influential ethical theories. Philosophers have held different views about what is good and bad, right and wrong and how we should lead our lives. Three prominent theories are consequentialism, deontology and virtue ethics, and each of these explains how the descriptive and normative enterprises relate to one another (§ 2). Next, we present these theories in more detail and situate them historically (§ 3). After that, we explain in what sense ethical reflection relates to our everyday lives, especially in light of the fact that academic debates can seem quite abstract and remote. Here, we also explain what distinguishes proper ethical reflection from moralizing (§ 4). We discuss a number of alternative approaches to ethics (§ 5) and end with a short reflection on the future of moral philosophy and the role we hope it will play with respect to the many societal challenges we are and will be facing (§ 6).

## 2. INTRODUCING THREE MAIN ETHICAL THEORIES

Let us return to the example of social media. Perhaps you tend to ignore or shrug off worries about privacy on social media or about governments and companies violating it. “I have nothing to hide, so be my guest,” you might think. Does that mean that you don’t care about privacy? Not necessarily. Imagine someone posting a funny but very embarrassing picture of a friend on social media without their permission. Or what about a future employer asking you for access to your medical and financial records, for example to be better able to evaluate your character? If you think there is something wrong in these cases, you *do* care about privacy after all. In a similar manner, you might not be

a vegetarian, but that does not mean that you do not care about animal welfare. After all, you probably criticize people who mistreat or neglect their pets and would be outraged if they were to eat them as well.

In such cases, philosophers typically ask “why?” What is the difference between a friend sharing personal information without consent and governments or big companies doing the same? What is the difference between treating pigs and cows as mere objects for consumption and mistreating or neglecting cats and dogs?<sup>1</sup> Which difference makes for a *moral* difference?

If we want to answer such questions, we need to systematically and critically reflect on what is good and bad, right and wrong, morally speaking. But doing so is not easy and requires that we first sort out what we value and care about and why. In contemporary moral philosophy, there are at least three influential theories or approaches—consequentialism, deontology and virtue ethics—that each take a different approach to what is good and bad, right and wrong, and how we should lead our lives.<sup>2</sup> As a result, these theories also have different ideas on why things like privacy and animal welfare matter to us and why they should.

The first main approach is *consequentialism*. According to this approach, consequences are the only thing that matters, morally speaking. To assess something, we *only* have to look at its consequences and not, for example, at the intentions of the agent. Think of money, which we can spend in a myriad of ways. To assess what is morally desirable, consequentialists say that we should simply do with money whatever generates the best possible consequences. *Utilitarianism* is a specific consequentialist theory that says that we should always try to generate the most utility, which can be understood hedonistically in terms of pleasure or more broadly in terms of well-being or happiness. What brings about the biggest sum of pleasure or happiness? Spending 50 euros on a nice dinner and a bottle of wine? Or spending the same amount on a charity that uses the money to relieve some of the dire needs of the global poor?

Or think about our activities on social media. Perhaps, utilitarians argue, sharing pictures is a way of multiplying happiness. The loss of privacy this entails, then, is only a small price to pay. Alternatively, utilitarians could object to posting a confidential confession of a friend on social media because it causes more harm (her painful embarrassment) than joy (your and other people’s pleasure).

- 
1. It is not our intention to suggest that considerations about privacy and animal welfare are on a par with each other in any respect. Instead, we want to stress that these are two very different issues that we might, at first sight, seem not to be concerned with.
  2. As we explain in § 5, there are many views critical of these most influential theories in ethics.



The second approach is *deontology* and focuses on rights that should be respected and duties that should be fulfilled, irrespective of their consequences. Deontologists typically make a distinction between what is good or valuable and the rights and duties that individuals have. Most deontologists would permit, for example, you're spending your money in whatever way you choose even when you choose to spend it on something that does not contribute to something good or valuable (it is your money, after all), as long as it does not violate someone else's rights (you cannot spend it to hire a hitman, for example). The same reasoning applies to social media: you may share your experiences with the world if you want to, but only to the extent that it does not violate someone's rights. Posting a confidential confession or a funny picture of someone without her knowing and consenting is a no-go for deontologists, regardless of how much pleasure it generates. This would be wrong, not because it would upset her (this is what utilitarians care about), but because you are violating her right to privacy, which entails the right to disclose information about her as she pleases.

The third approach, virtue ethics, thinks about morality in the light of virtues like honesty, trustworthiness and friendliness, more or less stable dispositions that people develop over time and that together make the person's life go well. According to virtue ethicists, the good, virtuous person leads a life that is worth living and that is fulfilled. The normative implication is clear: you should strive to become such a good, virtuous person. Posting embarrassing pictures of your friend or a confidential confession, a virtue ethicist could argue, reveals that you are not trustworthy. Such actions run counter to the kind of person you should try to become. When people succeed in becoming virtuous, they acquire what virtue ethicists refer to as "practical wisdom" (*phronesis*), the wisdom to act morally without the need to consciously reflect or deliberate. To find out which actions a virtue ethicist would prescribe (for example when it comes to spending money or behaving online), you may simply ask yourself what an exemplary, virtuous person would do. This strategy is the secular, ethical version of what Christians ask when in doubt: "what would Jesus do?"

Note that each of these theories explains what matters to us and why we care about certain things in different ways. This has normative consequences. If, for example, utilitarians are right and the reason why posting a friend's confidential confession on social media is wrong is that it causes harm, we can imagine scenarios in which this is actually not wrong: for example, when it does not harm your friend, because she does not care at all. Since nothing painful arises from your actions, consequentialists and utilitarians would not judge them to be wrong. According to deontologists and virtue ethicists, however, the wrongness of your actions does not consist in whatever harm they do and thus does not disappear if no harm was done. In this respect, different ethical theories provide

different descriptions and evaluations of our actions and hence have different consequences for what we should do.

In addition, each of these theories identifies a class of moral reasons that we sometimes appeal to when making decisions or justifying what we do. We often justify ourselves, for example, by claiming that no harm is done (a utilitarian response); or we believe fiercely that there are no conditions under which a certain act, say killing an innocent person, is allowed (a deontological response); or finally, we often praise someone for being courageous or kind (a virtue-ethical response). Hence, utilitarian, deontological and virtue-ethical theories, though elaborated on and examined by moral philosophers, are not far-fetched theoretical inventions unrelated to our everyday lives. In the next section, let us examine each of these theories a bit further.

### 3. FURTHER EXPLAINING THREE MAIN ETHICAL THEORIES

#### *3.1 Utilitarianism: Jeremy Bentham, John Stuart Mill and Peter Singer*

**T**he core of the utilitarian approach is consequentialist: consequences are what matters morally (about our actions, institutions, policies). There are no absolute values or rights (such as privacy) and no absolute principles or rules (such as never betray a friend) we need to respect. All we should do is promote the good or maximize good consequences; that is, we should promote or maximize 'utility'.

The good that has to be promoted, utility, can be and has been defined in quite a few different ways. It has been understood in terms of happiness and well-being, each of which has different possible interpretations and connotations as well. While some believe that your well-being and utility depend on whether your (subjective) desires are satisfied, others argue that there is an objective list of things that make you well off. Hedonic utilitarians in turn specify that the good that is to be maximized is happiness or pleasure. We should simply promote overall happiness, which is understood as maximizing the sum of pleasure minus pain. Jeremy Bentham (1789) was the first to articulate the well-known principle of utility—the greatest utility for the greatest number—which up to today serves as the core of utilitarianism. Something is more desirable, morally speaking, the more it promotes overall happiness.

Why is utilitarianism so influential? Happiness and pleasure are obviously desirable as they are things that all human beings strive for, while pain and harm are things all human beings aim for and have good reason to avoid. When we ask ourselves why we do the things we do and value the things we value, happiness is usually where we eventually end up. We value having nice dinners and meeting

up with friends because we enjoy them. Even things that are not enjoyable in themselves are ultimately motivated by happiness. Why study? To get a degree. Why get a degree? To get a job. Why get a job? To lead a comfortable life that you can enjoy. Why enjoy? As John Stuart Mill (1863: chapter 4) stresses, there is no further answer to this: “happiness is desirable, and the only thing desirable, as an end; all other things being only desirable as means to that end.”

The utilitarian idea that happiness and harm (to someone’s well-being) matter morally is widespread and uncontroversial in our current society. However, when this idea became popular in 18th-century England, it was quite radical and progressive. At the time, many things were considered morally wrong that did not necessarily harm anyone and that we have since ceased to see as wrong, such as homosexuality. Or take Mill again, who already claimed over 150 years ago that women should have equal rights on the basis of a utilitarian concern for overall happiness.<sup>3</sup>

Utilitarianism has an obvious ethical appeal since it identifies considerations to which we can all relate with ease. Since I care about my happiness and not being harmed, it is easy to understand that the same holds for you and all other beings like us. So, if we want to act ethically and treat others right, utilitarians argue, we need to look out for *their* interests as we tend to do for our own interests. In fact, because their happiness matters as much as our own, morally speaking, we need to maximize the total sum of happiness, in which each counts for one and none for more than one.

An important question here is whose interests matter and who belongs to the circle of moral concern (Singer 2011). If happiness and harm are the things that matter morally, then why stop with human beings? After all, a lot of animals can suffer too, and there is no good reason why their suffering should not matter morally either. As Peter Singer (1975) argues, why would the interests of sentient beings like cows, chickens, pigs and fish not matter as much as our own? As a result, we should stop eating meat and dairy products, since such diets contribute to animal suffering and are therefore immoral.

While the claim that happiness and suffering have moral relevance is uncontroversial, disagreement persists on what exactly they are and why we should attend to them. We often suffer pain and forgo instant pleasure to reach some higher goal, for example when we visit the dentist or when we refrain from a nice glass of wine before a sports match or job interview. Do all kinds of

---

3. He argued that equal rights for women would benefit both women (who are better off if they can freely choose what to do with their lives) and society (which will prosper thanks to contributions of women who can pursue their passions and develop their talents). He was among the first to make this case quite successfully.

suffering and happiness matter equally? Whereas Bentham (1789) understood happiness hedonistically as the mental state of feeling happy and argued that only quantity matters, Mill (1863: chapter 2) believed that quality matters as well. Enjoying a good novel is more worthwhile than eating a simple ice cream. If you agree with Mill, you could argue that the suffering and pleasure of human and non-human animals have different moral weight after all.

Utilitarians also disagree on whether we should maximize the consequences of every single action or rather stick to general rules or principles, such as “never betray a friend’s confidence.” So-called rule utilitarians argue that overall utility would be maximized if everyone followed such simple rules. According to so-called act utilitarians, every single action should maximize overall utility, but this has the problematic implication that it is perfectly fine to sacrifice some people’s interests for that greater good. If posting your drunk picture on social media generates more pleasure than pain, your privacy and embarrassment are outweighed by the fun other people have. In fact, an act utilitarian will claim that your friend (or even you) actually has a moral duty to make fun of you.

### *3.2 Deontology: Immanuel Kant*

**D**eontologists, in contrast to utilitarians, believe that there are some things that we are never allowed to do, such as violating someone’s privacy. We should follow absolute rules and principles, but not because of the happiness this brings about, as rule utilitarians would argue. According to deontologists, we have certain rights and corresponding duties simply in virtue of being humans. We recognize being bound by these moral precepts as this is, well, what morality demands from us.

Probably the best-known defender of deontology is the 18th-century philosopher Immanuel Kant. Kant believed, in contrast to many in his time, that in figuring out what to do morally, we should not turn to religion. Each and every one of us can, by virtue of our rationality, figure out what our moral duties are. Most of what we should or should not do is, as Kant puts it, regulated by so-called hypothetical imperatives. If we want to pass a course, we should study for the exam. The requirement (or imperative) to study is conditional (hypothetical) upon our desire to pass the course. Now, according to Kant, moral imperatives are, in contrast, unconditional or categorical in nature: they prescribe what we should do, regardless of our individual desires. Unlike other animals that pursue whatever their desires lead them to, humans possess the ability to reflect on those desires, whether they are worth pursuing and whether it is rational or moral to pursue them. Even if you know how much fun it would be to make fun of your friend online, you should refrain from it because you recognize that it is not the

right thing to do. Morality has nothing to do with pleasure or desires but with your capacity to reason.

Kant provides several formulations of the categorical imperative. The first is known as the “universalizability principle:” “Act only according to that maxim whereby you can at the same time will that it should become a universal law” (Kant 1993 [1785]: G 4:421). Imagine wondering whether meeting up with a friend is the right thing to do even though it means you have to break your promise to another, slightly less fun friend. According to Kant, you first need to identify the maxim, the general principle that underlies your action: “to break a promise.” Next, you should consider whether you could want everyone to act on that maxim. So you should ask what the world would look like if everyone broke their promises whenever something popped up they liked more. According to Kant, we cannot want this without contradiction. The very essence of promises is that we keep them, even when there is something we like more. Without this commitment, the whole idea of promises loses its meaning. So on purely rational grounds, you can come to understand the absolute moral principle that promises should be kept.

According to Kant’s second formulation of the categorical imperative, you should “act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end” (Kant 1993 [1785]: G 4:429). When you make fun of your friend online, you are treating her merely as a means (to generate pleasure or collect likes on your social media accounts) instead of respecting her as an end in itself. Both formulations of Kant’s categorical imperative serve as a kind of “litmus test” for morality. Whatever passes is morally permissible; whatever does not pass, you have a moral duty not to do.<sup>4</sup>

While there is quite a lot of discussion among contemporary deontologists, for example about which rights and duties we do in fact have or how they apply to concrete situations, they do share a common ground: namely, that rights and duties are the core of morality, not the utilitarian or consequentialist calculus of moral costs and benefits.

---

4. It is much more difficult to accommodate animal welfare within a Kantian framework although recently, Christine Korsgaard (2004), the most influential neo-Kantian, has argued in favor of animal welfare from a Kantian perspective.

### 3.3 Virtue ethics: Aristotle

The third main ethical theory is virtue ethics and dates back to the Greek philosopher Aristotle (fourth century BC). Aristotle (340 B.C. [1920]) argued that the “goodness” of something always relates to its purpose (*telos*) or function (*ergon*). Your ear is “good” if it enables you to hear. Lionel Messi is a “good” or “excellent” football player because he excels in those things (speed, accurate passing, quick thinking) that serve the purpose of the game. Messi has developed those capacities that enable him to do what he is meant to do, namely help his team win games. When acting out this purpose, Messi flourishes, doing what he is good at and what makes him happy.

Now, Aristotle argues in his *Nicomachean Ethics* (340 B.C. [1920]) that humans are “good,” morally speaking, if they excel in what makes them human, which in his view relates to their reason (*logos*). Since ethics is part of practical philosophy, it is not concerned with theoretical knowledge (knowledge for its own sake) but with practical knowledge (knowledge to do good and become a virtuous person). Just like Messi trained himself to know what to do on the football field, so should we, through practice and habituation, develop the practical wisdom (*phronesis*) that enables us to know what to do, morally speaking. The character (*ēthikē aretē*) that we should develop does not consist in speed or physical prowess but in moral virtues like courage, friendliness and temperance. And just as Messi thrives on the football field, moral exemplars thrive in life because they achieve *eudaimonia*, which can be translated as “happiness” or “human flourishing.”

Different virtue ethicists—Aristotle, but also contemporary authors such as Martha Nussbaum (1993) and Alisdair MacIntyre (1981)—may take different approaches within virtue ethics and disagree, for example, about which virtues are most relevant and how they relate to each other. However, all of them stress how crucial virtues are for answering what is perhaps the central ethical question: what is the good life? Because circumstances are always unique, virtue ethicists do not aim to prescribe specific actions or formulate universally applicable rules or principles. When you are a good friend, you know what the appropriate response is in a particular situation, for example when to post which pictures online (without having to weigh off utilities or check the categorical imperative).

Each virtue, Aristotle argues, is a mean between the extremes of deficiency and excess, each of which constitutes a vice. A courageous person balances between overconfidence and cowardice. When it comes to spending your money (on nice glasses of wine or poverty-alleviating charities), temperance forms the middle road between gluttony (spending all your money on indulgences) and total abstinence (never indulge in anything). The right path is that of moderation.

## 4. ETHICAL REFLECTION AND EVERYDAY LIFE

**B**y now, you should have some grip on what moral philosophy is and how it relates to our reasons for doing what we do and should do. Ethical reflection matters because moral considerations permeate our lives. Nevertheless, reading moral philosophy can be a frustrating experience.

On the one hand, philosophical discussions in ethics often employ far-fetched examples and thought experiments that seem completely unrelated to our everyday lives. Kant (1993 [1785]), for example, claimed that it is never permissible to lie, not even when a murderer knocks on your door and asks you to tell him where your friend, his intended victim, is. Or what about the famous trolley-dilemmas in which we are asked to imagine situations in which you either do nothing but allow five innocent people to die or do something that kills another innocent person (Foot 1967)? It is not the case that ethicists think these cases are likely to occur. Instead, they are discussing the strengths (and weaknesses) and (im)plausibility of competing ethical theories and their ability to guide us in our lives. How well do they connect with our moral intuitions, what are their implications, how sound is the advice they generate and how convincing are their arguments? This is a valuable exercise, especially when multiple incompatible theories make intuitive sense. However, when you turn to moral philosophy for practical advice, such discussions can be disappointing and frustrating.

On the other hand, philosophers who do have explicit practical advice and have firm ideas on what you should or should not do are often perceived as preachy and moralistic. They seem to take the moral high ground with advice that is often difficult to follow or that takes all the fun out of our lives. This reproach can apply to utilitarians (who tell you to stop spending your money on frivolous things and help eradicate global poverty instead), deontologists (who tell you to ignore the consequences of your actions) and virtue ethicists (who tell you to stop slacking and start developing your character). This could lead you to respond in several ways: “Who does this philosopher think she is, telling me what to do?” or: “What fun is left in leading an ethical life? Life is too short to worry about all those things!”

As mentioned, we take ethical theories to be highly systematic and abstract elaborations that are grounded in everyday moral intuitions and considerations. Their often radical conclusions result from the willingness of philosophers to think things through and go where the argument leads them.<sup>5</sup> However, it is always open to each of us to explain what, if anything, is wrong with the view

---

5. If you really think that cows, pigs and chicken can suffer and that this matters morally speaking, then the argument for veganism is actually quite straightforward.

or argument in question. This is why we have emphasized that ethics concerns us all and why we should be willing to think critically about the moral intuitions, beliefs and claims of ourselves, others and society. As in other philosophical disciplines, the appropriate attitude when we engage in such reflection is one of humility (“this is what I think but I might be wrong”) and openness (“others might be right”). As such, the kind of moralizing that ethicists are sometimes accused of embodies the exact opposite attitude, namely one of arrogance, overconfidence (*hubris*) and closed-mindedness.

The criticism of ethics as being either too far removed from everyday life or too moralistic has also been voiced within philosophy. Some philosophers, for example, have argued that these theories fail to capture the nature of everyday moral reasoning. How often do you pause to consider the maxim of your action and whether you can universalize it? Have you ever thought of trying to maximize the sum of pleasure and pain of all sentient beings in the world and thus live up to the extreme but perfectly consistent outcome of the utilitarian creed that all interests have moral weight?<sup>6</sup>

Others have pointed out that the prescriptions of ethical theories are often overdemanding and psychologically unrealistic. To what extent are ordinary human beings like ourselves actually able to follow the advice that follows from these theories? Again, think of Kant’s advice to never, under any circumstance, tell a lie. Or Singer’s utilitarian requirement to donate most of our money to charities that reduce global poverty (Singer 1972). Who can live up to this?

The final but perhaps most worrying criticism from within the discipline is that the very way in which ethics is practiced in mainstream theories is wrongheaded at its core. This has led to several alternative or non-mainstream approaches, to which we turn next.

---

6. The English philosopher George Edward Moore (1903) famously argued that utilitarians and Kantians misrepresent the moral domain. He pointed out that we typically know what things are morally good or bad without engaging in utilitarian calculations or Kantian exercises. Surely, friendship, love and trust, and integrity are good, and harm, dishonesty and betrayal are bad. Interestingly, while Moore’s observation seems to simplify things, it actually complicated them tremendously. How is it that we know these obvious moral truths? Do we have a special sense or intuition to trace them? And what is the status of this knowledge and of these truths? Do ethical values somehow exist in the same way as other things exist? These questions take us into the domain of meta-ethics and beyond the purposes of this chapter.



## 5. ALTERNATIVE APPROACHES IN ETHICS

**D**eontological and consequentialist theories are characterized by a specific way of thinking that involves an impersonal point of view and strict principles or rules that are translated into action-guiding prescriptions.<sup>7</sup> According to some philosophers, this fails to capture the nuances and complexities of our lives and the enormous role that our love (rather than duty) for particular things and people (rather than general principles and rules) plays in this respect (Williams 1981, Frankfurt 1982). In addition, impartial and universalizing approaches to ethics arguably neglect the extent to which our lives are shaped by our specific relationships with others and our individual, social and cultural histories (MacIntyre 1981, Taylor 1989).

According to some more radical critics, such as Margaret Walker (2007), this failure and neglect has led to a biased understanding of what morality and moral agency are all about. While navigating different responsibilities in life and taking care of people around you is clearly of moral relevance, it is hard to capture in the all-or-nothing vocabulary of rules, principles, duties and rights.<sup>8</sup> The traditional emphasis on this vocabulary has already, for example, led to the mistaken idea that females are less morally developed than males.<sup>9</sup> This criticism has been central to the development of an “ethic of care” (Gilligan 1982, Tronto 1993). According to feminist philosophers, more generally, we cannot do proper moral philosophy without attending to and being critical of existing and widespread power imbalances in our society. For example, traditional ethical theories pay little attention to the experience of people who do not feel at ease with the moral views of the society they live in because they belong to a repressed minority group. Recently, feminist philosophers have analyzed exactly those experiences in terms of, for example, moral disorientation (Harbin 2018) and affective injustice (Srinivasan 2018) and, hotly debated in the past decade, epistemic injustice (Fricker 2007).

By now, the multitude of approaches in ethics provide a more nuanced picture of the moral domain, with interpersonal relationships and the social dimension of our living together taking up a more central place. Empirical work in the

- 
7. Virtue theory is considered an exception to this, and is therefore often employed as a criticism of an alternative to consequentialism and deontology.
  8. “Navigating responsibilities” means making the best of one’s situation, for example by trying to take care of those around you to the best of your abilities. This requires paying attention to the details of people’s specific situations rather than abstracting from them and trying to pinpoint what general rights or duties are at stake here.
  9. Psychologist Laurence Kohlberg (1981), for example, found a difference in moral development between women and men based on his Kantian-inspired ideas on morality. Carol Gilligan (1982) developed her alternative “ethics of care” in response to his ideas.

behavioral sciences has also caused a surge of interest in the moral motivations behind everyday decisions and has prompted more realistic accounts of our moral psychology and agency. In the meantime, feminists and non-mainstream ethicists continue to offer highly critical analyses of our societies that, according to some of them, are fundamentally racist, sexist and ableist (discriminating on the basis of ability).

While this chapter is too short to do justice to these critical perspectives, we think and hope that they will continue to change the way ethicists and ordinary people think about morality. In fact, many of the challenges we face in today's societies go beyond individual responsibilities and require more complex answers than the more traditional ways of thinking are able to provide. Asking what we, as individual agents, ought to do (donate more money to charities, change our diets, post pictures on social media, etc.) seems to ignore the frameworks that provide and structure these different options. Challenging those frameworks themselves can mean that we take a critical look at the role of companies who develop these technologies, the capitalist system they operate in and the impact that technologies like social media have on our relations with one another, ourselves and the world. In our concluding section, we go into this in a bit more detail.

## 6. WHERE ARE WE HEADING?

So what about the future of moral philosophy? Which topics will occupy moral philosophers of the future? First and foremost, the rapid advancement of technological innovation will inevitably continue to raise ethical issues. Of course, some of these can be insightfully examined from the perspective of the traditional accounts.<sup>10</sup> However, evolutions such as the advent of artificial intelligence, the quickly increasing capacities of machines for self-learning and deep learning, the possibilities of virtual and augmented reality, the advantages and threats of big data and the permeation of social media into our lives all confront us with ethical issues that go far beyond our individual responsibilities. These technologies have social, cultural and political effects that are still difficult to envision, even for those most knowledgeable about them. Rather than merely providing us with ever-more sophisticated tools and options to lead our lives, they also change the world we live in and the ways in which we relate to that

---

10. Take the challenges we face when it comes to the use of personal data based on social media. As various scandals have shown, people do care about their personal data being protected better and having more control over what happens to it. While this is the basis of the new EU law on privacy and data protection (GDPR), it is important to acknowledge that such laws will not suffice. We should all become aware of the dangers involved when we allow (often huge) companies to access and acquire our digital traces and use them to their advantage. When trying to behave responsibly online, we can think about what would harm people, what would violate their rights or simply what a virtuous person would do.

world, to ourselves and to others.<sup>11</sup> With a smartphone, for example, you navigate your social and physical surroundings in completely different ways from before (or from someone without a smartphone).

Other issues, less directly related to technology, will remain or become relevant as well. How should we proceed as climate change unfolds (remember that it is already too late to avoid or stop it)? How can we avoid the polarization that permeates a lot of contemporary societies? How can we make sure that the interests of the poor are not overruled by those of the rich? What kind of duties do companies have toward their consumers and employees?

While some of these issues will be about life and death, peace and war, others will be about the seemingly low-stakes decisions that make up everyday life. In any case, we hope that we have made it clear how ethical reflection might help in tackling the challenges we are facing, small and big, at both the individual and the systemic level. Even though there is no pause button in real life, taking the time to thoroughly reflect on and discuss with one another who we are and what our values entail is a necessary step in leading better lives and designing better institutions.

---

11. The naive conception of technology as adding options and tools to our lives that we can evaluate in isolation has been criticized by philosophers such as Martin Heidegger (1954 [1977]), Hannah Arendt (1958) and Max Horkheimer and Theodor Adorno (1947 [2002]).

## BIBLIOGRAPHY

- Arendt, H. (1958). *The human condition*. University of Chicago Press.
- Aristotle. (1920). *Oxford classical texts: Aristotelis: Ethica nicomachea* (I. Bywater, Ed.). Oxford University Press. <https://doi.org/10.1093/actrade/9780198145110.book.1>
- Bentham, J. (1789). An introduction to the principles of morals and legislation. In *The collected works of jeremy bentham: An introduction to the principles of morals and legislation*. Oxford University Press.  
<https://doi.org/10.1093/oseo/instance.00077240>
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect. In *Virtues and vices* (Vol. 5, pp. 19–32). Oxford University Press.  
<https://doi.org/10.1093/0199252866.003.0002>
- Frankfurt, H. (1982). The importance of what we care about. *Synthese*, 53(2), 257–272.  
<https://doi.org/10.1007/bf00484902>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Gilligan, C. (1982). *A different voice*. Harvard University Press.
- Harbin, A. (2016). Disorientation and moral life. *Studies in Feminist Philosophy Series*.
- Heidegger, M. (1977). The Question Concerning Technology. In W. Lovitt (Trans.), *The Question Concerning Technology and Other Essays* (pp. 3–35). Harper & Row.  
<https://doi.org/10.5040/9781472547873.ch-002> (Original work published 1954)
- Horkheimer, M., & Adorno, T. (2002). *Dialectic of enlightenment: Philosophical fragments* (G. S. Noerr, Ed.; E. Jephcott, Trans.). Stanford University Press. (Original work published 1947)
- Kant, I. (1993). *Grounding for the Metaphysics of Morals* (J. W. Ellington, Trans.). Hackett Publishing Company. (Original work published 1785)
- Kohlberg, L. (1981). The philosophy of moral development. In *Essays on moral development: Vol. I*. Harper & Row.
- Korsgaard, C. (2004). Fellow creatures: Kantian ethics and our duties to animals. In B. Peterson (Ed.), *The tanner lectures on human values* (Vol. 25). University of Utah Press.
- Macintyre, A. (1981). *After virtue: A study in moral theory*. University of Notre Dame Press.
- Mill, J. S. (1870). *The subjection of women*. Longmans, Green, Reader, and Dyer.  
<https://archive.org/details/subjectionofwom00mill>
- Mill, J. S. (2017). *Utilitarianism*. Parker, Son, and Bourn.  
<https://archive.org/details/a592840000milluoft> (Original work published 1863)
- Moore, G. E. (1903). *Principia Ethice*. Cambridge University Press.
- Nussbaum, M. (1993). Non-Relative Virtues: An Aristotelian Approach. In M. Nussbaum & A. Sen (Eds.), *The Quality of Life* (pp. 242–269). Oxford University Press.  
<https://doi.org/10.1093/0198287976.003.0019>
- Singer, P. (1972). *Famine, Affluence, and Morality*. 1(3), 29–243.  
<https://doi.org/10.4324/9780203723746-85>

- Singer, P. (1975). *Animal liberation: A new ethic for our treatment of animals*.
- Singer, P. (2011). *The expanding circle*. Princeton University Press.  
<https://doi.org/10.1515/9781400838431>
- Srinivasan, A. (2018). The aptness of anger. *Journal of Political Philosophy*, 26(2), 123–144. <https://doi.org/10.1111/jopp.12130>
- Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Harvard University Press.
- Tronto, J. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge.
- Walker, M. U. (2007). *Moral understandings: A feminist study in ethics*. Oxford University Press.
- Williams, B. (1981). *Moral luck*. Cambridge University Press. ONLINE VIDEOS.

## ONLINE VIDEOS

<https://www.khanacademy.org/partner-content/wi-phi/wiphi-value-theory#wiphi-ethics>: This is an entire series on ethics, including a video on consequentialism by Julia Driver and a three-part introduction on utilitarianism by Julia Markovits.

<https://www.khanacademy.org/partner-content/wi-phi/wiphi-value-theory#wiphi-good-life>: This is a short collection of four videos on the good life, including Aristotle's and Kant's views on this (next to Plato's and Nietzsche's).

<https://www.youtube.com/playlist?list=PL8dPuuaLjXtNgK6MZucdYldNkMybYIHKR>: This "crash course philosophy" (by CrashCourse) introduces basically all philosophical disciplines and has a section on "fundamentals of ethics." Besides utilitarianism, deontology and virtue ethics, this has videos on some slightly less influential theories, such as natural law theory and contractarianism.

VIRGINIA HELD



## VIRGINIA HELD (2006)

### *THE ETHICS OF CARE: PERSONAL, POLITICAL, AND GLOBAL*

---

**A**ccording to Virginia Held (°1929), we all stand in particular relationships of care to others. In those relations, our emotional responses to one another have an important role to play as well as the particular claims of the people with whom we have these relationships: our children, parents, neighbors, and so on. When talking ethics, we should not strive for impartiality, abstract away from those relationships and neglect the responsibilities these relationships bring. The values that regulate our personal relationships, such as care, trust and solidarity, fit with and are as least as important to the moral domain as the more traditional values of equality, justice and rights. Conventional ethical theories such as utilitarianism and deontology are inadequate, Held argues in her recent but already influential book, not only when it comes to guiding our personal lives, but also when it comes to dealing with the political and global challenges we are currently facing. She argues that an ethics of care is much better equipped to deal with contemporary societal and political problems such as the role of markets and the power and violence that characterize all relations, including those at the global level. If we want people and states across the world to live in peace, respect human rights and care for their environments, Held stresses that we need to globalize caring relations.

---

Thijs Lijster

**Imagined and  
Contested  
Communities**



# IMAGINED AND CONTESTED COMMUNITIES

## *An Introduction to Philosophy of Culture*

Thijs Lijster<sup>1</sup>  
t.e.lijster@rug.nl

### 1. INTRODUCTION

Soon after the Argentine-born Máxima Zorreguieta became princess—and therewith, soon-to-be-queen of the Netherlands—she gave a speech in which she denied the existence of “*de Nederlander*” (*the Dutch person*). In her view, the Dutch population was far too diverse to be encompassed within one single image. A storm of outrage followed: critics accused her of being too ‘politically correct’ and labeled her a cultural relativist. Of course, she was right: there is no ‘typical’ Dutch person, and nor is there an average European, Asian or African. Still, that does not mean that we cannot speak of or think about Dutch culture, about differences between Asian and African cultures, or about any other culture in a meaningful way. It does, however, mean that culture is a complex, difficult to pin down, and contested phenomenon.

What is culture? What are its origins? What constitutes a culture, and how is culture distributed, sustained and transformed? These are some of the central questions within the philosophy of culture. Although its research subject—human culture—is as old as humankind itself, the philosophy of culture is among

---

1. I want to thank the friends and colleagues who have commented on earlier versions of this chapter, especially René Boomkens, Bart Engelen, Martin van Hees, Lodi Nauta, Maureen Sie, and Rosie Taekema.

the youngest branches on the tree of philosophy. This is probably because for a very long time in human history, culture was not conceived as a problem. Although someone like Michel de Montaigne had already questioned the extent to which European culture was superior to the customs of the people living in the recently 'discovered' Americas, it was not until the late eighteenth century that Johann Gottfried Herder argued how each culture is unique, and that "nothing is more deceptive than the application [of a single concept of culture, TL] to entire peoples and eras" (Herder quoted in Konersmann 2003, 12).

Because of this, the questions that concern the philosophy of culture are constantly changing. Not only according to scientific insights in related academic fields (such as anthropology, history, sociology, and media studies, as well as other subdisciplines of philosophy such as political philosophy and the philosophy of technology), but also according to public debates on culture and identity. Perhaps more than any other branch, then, the philosophy of culture fits Hegel's definition of philosophy as "its own time comprehended in thoughts" (Hegel 1991/1821, 21).

In this chapter, I will first discuss the concept—or rather, concepts—of culture. How we define culture is very often not value-neutral, and also hangs together with *who* is defining or describing whose culture. Hence, we will look next at the relationship between culture and power. After determining culture as a primarily collective practice (rather than as a set of objects), I will look at how culture is transmitted and distributed through media. Finally, I will discuss what culture, and the philosophical study of culture, mean in a globalized world in which different cultures are increasingly confronted with each other—sometimes mixing and sometimes clashing.

## 2. DEFINING CULTURE

'Culture' is one of those terms we use so often that we hardly think about what it means. We speak of 'local cultures' and 'popular culture'; we have a 'ministry of culture', go on 'cultural vacations', and we even wage 'culture wars'. But is there a common meaning to be discovered across all these different ways of using the concept?

The British philosopher Raymond Williams famously defined three distinct uses of the concept of culture (1976). The first, is the idea of culture as 'civilization', as it was commonly used in the age of the enlightenment, and which is closely connected to ideas of progress and universalism. When thus applied, one usually presumes that only some groups of people 'have' culture, while others are considered 'primitive'. This also means that the primitive groups are supposed to, eventually, live up to our standards. The second, is culture as defined by the

cultural critic Matthew Arnold, as the “pursuit of our total perfection by means of getting to know [...] the best which has been thought and said in the world” (1869). This account is usually applied to the arts, and culture is then seen as a synonym of ‘high’ art as represented by the canon of the great works of tradition. Finally, we have the broad anthropological use of the concept of culture as being the culture of a certain group. In this third use, culture is understood as collective practices of meaning-making. Although each of these uses still come up regularly in contemporary debates, and will also return in what follows in this chapter, the latter, broad concept of culture clearly predominates today. As Williams put it in a different essay, “Culture is ordinary” (1977); culture is not something reserved for the higher social classes or for western civilization, but is simply people’s everyday customs and practices, and the significance they attach to them.

Culture has often been considered as the defining feature of humankind. Humans are, to borrow a phrase by Helmuth Plessner, “by nature artificial” (2019/1975), meaning that we can only survive by actively intervening in our natural environment: since we don’t have fur, we need clothes and shelter; since we don’t have claws, we need tools. Furthermore, we need to teach these practices to the next generation in order for them to survive. Hence, signs and language play an important role in culture (as we will see later on when discussing media). Still, while some species of animals also use tools and build shelters, what actually sets humans apart is our reflection on such practices and the meanings we ascribe to them. According to Martin Heidegger this has to do with the awareness of our mortality (1996/1927): the realization of the finitude of our lives on this planet causes the desire to somehow make sense of, and give meaning to our existence—through rituals, stories or myths, and works of art. This meaning can only be granted by, and in relationship with others (other people, as well as animals, objects, etc.). Therefore, culture always has a social and relational dimension; one cannot ‘do’ culture on one’s own.

While culture thus belongs to the very nature of humankind, the varieties and diversity of different cultures is endless. And, as we learn more from other cultures, we might begin to question our own. The British-Ghanaian philosopher Kwame Anthony Appiah argues that culture is closely connected to identity—that is, to the ways in which we consider ourselves and others (2018). Appiah describes (cultural) identity on the basis of three characteristics. First and foremost, identity is a label with which we tend to classify ourselves and other people, along lines of creed, country, colour, class, gender, and so forth (as in: I am a catholic, a Frenchman, et cetera). Secondly, such labels provide guidelines and orientation for people in their actions, values and beliefs (e.g. ‘as a European, I am tolerant’, or ‘as an American, I am proud of my country’). Thirdly, identity can also determine the ways in which others think of or treat you, in either a positive or negative sense; for instance, in the form of prejudice or stereotyping.

History is, unfortunately, full of examples of crimes that were committed against people simply because they belonged to a different culture. But, Appiah states, cultural identities are never strictly defined—their boundaries are never fixed but are always in motion, because history itself is in motion. They are constructions which, nevertheless, have an enormous power to connect people to, or alienate them from each other; hence he calls them, ‘the lies that bind’.

Cultural identities are ‘fluid’, which means that they do not have a fixed, deeper core or essence; they are not solid but ‘liquid’ (to follow the imagery of Zygmunt Bauman, 2000). On the basis of this, it is sometimes suggested that we therefore have a great freedom to construct, or even to choose, our identity at will, simply by behaving or dressing differently. Appiah points out, however, that identity has both a subjective and an objective dimension, and the latter is sturdier than notions of ‘fluidity’ or ‘liquidity’ imply. After all, in many cases, we *are* identified in a certain way—categorized according to certain existing cultural and social parameters or norms. This implies that our identities are not a merely subjective, individual matter, and hence not, or at least not entirely, for us to ‘construct’; this is rather a social, intersubjective process. Our identities always, to a certain extent, conform to existing cultural and social norms. To further analyze that process, it is important to look at the element of *power* that is part of any discussion on culture.

### 3. CULTURE AND POWER

It has already been mentioned that an important function of culture is a socializing one: in and through culture, we learn the ways of acting and behaving, and the beliefs, symbols and norms of the society we are born into. Thus, culture contributes to the integration of the individual in a specific social, political and economic order. This raises the question of who determines these norms and beliefs. Are we all equally capable of constituting the culture that we are part of? Historical and contemporary struggles of suppressed minority groups demanding the right to freely express and transmit their culture, are clear evidence that we are not all equally involved, and that, in fact, there have always been groups in society whose culture is considered to be more valid and valuable than that of others.

Karl Marx is one of the main sources of inspiration for modern and contemporary cultural theory, and one of the first who pointed to this close connection between culture and power. This is ironic, since Marx criticized such emphasis on the power of ideas from predecessors like Herder and Hegel. While Hegel considered human history as the progression of human self-consciousness, Marx rejected this ‘idealism’ (i.e. the belief that ideas are the driving force of history) and replaced it with his ‘historical materialism’. In his view, the material

and economic circumstances of a society determine human consciousness. This view is elaborated in the so-called 'base-superstructure' model, according to which the socio-economic 'base' determines the ideological 'superstructure' (1978/1859). What Marx calls the 'base' consists of the means of production (tools, technologies and materials, but also human labour power) and relations of production: social relations between the class that owns the means of production and the one that does not—the latter of which, therefore, has to sell their labour power to the one with the ownership. According to Marx, history is a continuous 'class struggle' in which a certain class (in his age, the industrial proletariat) is exploited by the owners of the means of production, until they can no longer take it and start a revolution.

By 'superstructure', he meant all spiritual products of man: religion, art, science, but also law and politics (so, actually, the entire culture). In Marx's metaphor, the superstructure rests on the base like a house on its foundations; in other words, culture depends on the economic system. This implies that culture is a form of ideology, as a reflection of existing class power, as well as a legitimization thereof. Take religion as an example: for a long time, Christianity sustained existing relations of power because it made the oppressed class believe that God created the social order as it should be—if you do your duty and accept your suffering, you will be rewarded in the afterlife. Marx famously called religion the 'opium of the people', because it keeps the masses peaceful and quiet.

There are at least two problems with this view on the relationship between culture and power, both of which have been pointed out by the twentieth-century French philosopher Michel Foucault (1982). Firstly, the power struggle that Marx is talking about is not only taking place between two main social classes, but also takes place within classes, between different groups, and on different socio-cultural levels. During the early twentieth-century struggles for women's voting rights, for instance, women from different classes stood side by side, since women were being suppressed within both the upper and the lower classes. The same goes for other forms of exclusion (for example, on the basis of ethnicity or sexual orientation) which cannot be caught in Marx's binary logic of class struggle. Secondly, Marx talks about power as the repression, by means of laws and/or violence, of the otherwise 'free' individual. For Foucault, however, there are far subtler, as well as more effective forms of power, in the shape of caring, monitoring, and thus disciplining. These he calls 'pastoral power' because the way that modern institutions rule over us best resembles the way a pastor keeps watch over his flock; to be able to guide everyone on the path to salvation, the pastor has to *know* what is going on within the congregation, the sins and secret desires of each of the members. This form of power is thus based not on violence, but on knowledge and specific uses of language: what Foucault calls *discourse*. Two fields of cultural theory that have been very much influenced by these ideas

are gender studies and postcolonial studies, which I will briefly discuss to further elucidate the relationships between culture and power.

The French feminist philosopher Simone de Beauvoir famously stated that one is not born a woman, but rather becomes one (2015/1949). The identification by others, discussed earlier, starts at birth, if not already in the womb (“is it a boy or a girl?”). This seems like a mere fact of nature, but in fact is a label that comes with all kinds of connotations. Gender creates expectations: e.g. boys don’t cry, and ‘girls just want to have fun’. According to Hélène Cixous and Catherine Clément (1986), such expectations are deeply ingrained in our culture, and are even part of our language. Gender differences are an element of a linguistic system of binaries (between culture and nature, mind and body, subject and object, reason and passion, active and passive, etc.) in which one tends to be considered as superior to the other. Men are supposed to be rational and proactive; women were traditionally conceived of as passive, emotional (and hence irrational) and weak. To consider such differences as biologically determined, of course, serves the existing power relations between genders (e.g. “women are by nature caring creatures that simply don’t want to have be in a leading position, but rather want to stay at home with the kids”). Conversely, by studying such differences as cultural phenomena (which is not the same as denying the existence of gender differences altogether), one opens up the possibility to change gender roles, a possibility which is absent if one takes a more naturalistic perspective on identity.

Closely related to these issues are the ways in which minorities and non-western cultures are represented by western culture and media. This is one of the subjects of postcolonial studies—an academic field that studies the history and legacy of western colonialism. Colonialism was never a mere political or economic matter, but also implied cultural hegemony—that is, the power to describe cultures. A classic in this field is *Orientalism* (2003/1978) by the Palestinian-American scholar Edward W. Said, which is an analysis of the ways in which European colonial reign was preceded and legitimized by the production of a body of knowledge and imagery of Eastern countries and cultures. By looking at, for instance, 18th and 19th century literature and visual art—such as Flaubert’s letters from Egypt or Delacroix’s paintings—Said shows how a picture of the Orient was created as being exotic, mysterious, and sensuous but also irrational, backwards and barbarian. At the same time, and by creating this image, the West started to define itself in the opposite way, as innovative, modern, advanced, and rational. Said writes: “Therefore as much as the West itself, the Orient is an idea that has a history and a tradition of thought, imagery, and vocabulary that have given it reality and presence in and for the West.” (2003, 5) This idea pervades up until this day—first of all, in representations within popular culture (for instance, the way the Orient is portrayed in Disney’s *Alladin*, or the way Hollywood action movies tend to portray Arabs as terrorists), but, moreover, it also still functions

as political legitimization of more recent Western interventions in Iraq and Afghanistan.

This once again shows that the discourse on culture—and our knowledge and image of it—is never value-neutral, but is thoroughly intertwined with relations of power. Part of the research within the philosophy of culture is to understand and map such power relations. These are never as one-dimensional or binary as Marx thought (as simply being the power of one class over the other). In Said's analysis of Orientalism, too, other differences are obscured (such as the fact that the kind of western fantasy of the East he describes was primarily a male one). The black feminist scholar Kimberlé Crenshaw (1989) coined the term 'intersectionality' to point out how certain groups, such as black women, are dealing with a particular form of suppression that is not sufficiently addressed or recognized by (in this case) either (white) feminists nor by the Black Power movements at that moment (which focused mainly on male issues). Such groups are on the 'intersection' of different forms of discrimination and oppression, to which we could also add (dis)ability, sexual orientation, age, and so on. How minorities are represented in media—traditionally in painting or literature, and today in movies, popular music and television, forms an important part of such forms of oppression.

#### 4. CULTURE AND MEDIA

Imagine that someone tells you they're going on a 'cultural vacation' to Paris. You would probably expect that person to go see the Louvre, the Notre Dame, and perhaps go to the opera or a concert rather than, say, go shopping, drinking or visiting Disneyland.

Here, we have the concept of culture as 'high art' that we mentioned in the beginning. Throughout modern history, cultural institutions such as museums, theatres and cultural heritage sites, have played a decisive role in the constitutive process of (national) cultures, in constructing cultural hierarchies through processes of canonization. Only those objects and ideas that were considered to be valuable enough to be preserved and remembered, were placed in a museum or library, or were taught at universities. One could also say that these institutions were the media through which culture was disseminated to the wider population and transmitted to future generations.

While the production, distribution and consumption of this kind of culture was, for a long time, the privilege of the happy few, the late 19th century saw the dawn of *mass culture* made possible by technological innovations such as photography, film, radio and the gramophone. Cultural philosophers of the so-called 'Frankfurt School' were among the first to seriously study this mass culture, although they evaluated it in various ways.

Walter Benjamin, in his famous essay *The Work of Art in the Age of its Technological Reproducibility* (2008/1936), asks what the consequences are of (in his time) new media for our experience of art—most notably, visual art. Images generally used to have a religious meaning; think of religious icons, the statues of Greek gods, or the images of Mother Mary or the saints in the Catholic Church. This ‘cultic value’, as Benjamin calls it, still resounds in our respect for authentic masterpieces of art, and the fact that thousands of people everyday want to catch a glimpse of da Vinci’s *real* Mona Lisa or Rembrandt’s *Night Watch*.

But, while the creation of images used to be a time-consuming enterprise, photography, for the first time, made it possible to make an image in the blink of an eye, and to reproduce it limitlessly. Moreover, in the case of photography and film, it no longer makes sense to speak of a singular ‘original’ as opposed to its copies: the same work of art can be seen anywhere at the same time. In other words, as a consequence of these means of technological reproduction, images now surround us in our everyday lives: on posters and advertisements, in magazines and movies, and, more recently, on tablets and smartphones. This drastically alters our traditional attitude towards images—they lose their religious significance (what Benjamin calls their ‘aura’, 2008/1936) and become something mundane for us. According to Benjamin, this process had a potentially democratic effect, since the mass of the people are now themselves involved in the creation and distribution of images, rather than merely bowing before the images handed over to them by tradition.

His friend and colleague Theodor W. Adorno, however, thought otherwise. In Adorno’s view, mass culture was not a form of democratization, but rather ‘mass deception’. It did not emerge spontaneously from the masses—from the people’s own needs and desires—but rather was imposed upon them from above by media companies that served only one purpose: profit. Therefore, he preferred to speak of the ‘culture industry’ rather than of mass culture (Horkheimer and Adorno 2000/1947). This term also refers to a specific mode of the production of culture which resembles the industrial production of commodities. Principles of standardization and efficiency had now also started to dominate the realm of culture—each production company wanting to make a ‘hit’ or a ‘blockbuster’—thus seriously reducing the creativity and spontaneity of both the artist and the public.

Adorno’s view still resounds today in often-heard complaints about the uniformity of Hollywood movies or thirteen-in-a-dozen pop stars. However, Adorno’s point was not that the mass audience lacked good taste; rather, he saw the danger of a possible dwindling of critical thought. The freedom and creativity originally offered by culture also allowed it to be a realm of critical reflection on society. Art, in his view, always shows us different worlds—thus, critically questioning the existing one. As soon as culture exists merely in the service of profit



maximization, Adorno argued, it tends to become standardized, and thus loses this critical capability, and only further affirms the status quo. For Adorno, who, as a Jew, had to flee Europe in the early 1930s, it was of utmost importance that the masses have the capacity to critically reflect on the societal system that they are part of.

However, Adorno's negative take on mass culture has also been criticized by cultural philosophers and media scholars. First of all (and especially since the 1960s), mass culture has not been as uniform as Adorno claimed it to be. On the contrary, popular culture turned out to be an important carrier of a protest generation and subculture that, for instance, protested against war, racial oppression, and inequality (think of protest songs by Bob Dylan, or movies like *Easy Rider*). Rather than being a monolithic ideological apparatus, contemporary cultural philosophers and theorists consider mass culture as an arena where different world views and interests are competing with each other. Moreover, the profit motive need not undermine the critical role that culture can have. A movie like *Black Panther* was both a critique of colonialism and white supremacy, as well as a savvy superhero blockbuster.

Secondly, on the consumer side, the audience might not be as mindless as Adorno envisioned. Cultural theorist Stuart Hall (1993) argued that cultural productions are inherently 'polysemic', meaning that they can be interpreted in different ways. He described this as a process of *encoding and decoding*, in which a preferred meaning is 'coded' into a cultural product and subsequently interpreted by the audience. This 'decoding' can happen in line with the dominant and intended values but the audience can also take in an oppositional position. While the preferred meaning of Batman is that of a caped crusader fighting for justice in a crime-infested city, an oppositional reading might consider him a fascist billionaire who is merely maintaining the domination of his class.

Media, however, are not merely a means of distributing culture; they have the power to change culture as well. As philosopher Marshall McLuhan famously said, the medium *is* the message (1964)—by which he meant that, with the introduction of new media, our culture will change as well. For example, the world with television or internet is radically different from the world without, and, if Twitter allows you to only use 280 characters, this will alter the nature of your messages. In our media-saturated society, advertisements, movies and online content have a huge influence on how we think about, say, beauty or love; and, in politics, for instance, the 'image' often seems to be more important than the content of issues at hand. The images that surround us in our everyday lives are therefore not mere representations, but also determine the ways in which we view the world. Jean Baudrillard said that these images—'simulacra' as he called them—become more real than the reality they are supposed to represent; they become a 'hyperreality' (1995/1983).

Many consider this to be a disquieting development. If images become more real than the things they represent, how can we still distinguish the fake from the real, and truth from lies? Are we indeed living in a 'post-truth' condition, as is sometimes suggested—that is, a situation wherein the truth gets buried under a cacophony of opinions and worldviews? Bruno Latour (2004) argues, however, that there is a difference between saying that truth is context-dependent and saying that there is no truth whatsoever. The truth was never 'out there', waiting to be discovered—we always had to 'make' it ourselves. And this did not happen automatically: one needs to research, negotiate, translate, convince, etc. What the discussions on 'post-truth' do show, is that truth depends upon a common agreement on how to determine it, and a basic trust in certain institutions (such as universities and news media) to honour this common agreement. That this agreement and trust are increasingly lacking, seems to be a negative side-effect of the rise of social media.

## 5. GLOBALIZATION

Images travel fast in this digital age. Cartoons depicting the prophet Mohamed published in a Danish newspaper in 2005, for instance, led to violent demonstrations in Muslim countries, a ban on Danish products, and even an attack on one of the cartoonists. But it is not only images that travel faster: also information, commodities, capital, and people. For several philosophers, such as Paul Virilio (2006) and Hartmut Rosa (2013), modernity itself can be characterized as a process of increasing acceleration, leading to a compression of space and time. As soon as one can be anywhere in no time, both time and space tend to lose their significance. According to some philosophers, this process of globalization will lead to an overall homogenization of culture. Western, and particularly American, mass culture dominates the world: we are all wearing the same jeans, listening to the same music and eating the same hamburgers. Will the traditional diversity of cultures soon cease to exist because of such a "McDonaldization" of our world?

The Indian-American scientist Arjun Appadurai denies this. According to Appadurai (1996), the driving force behind the globalization process is *imagination*. This imagination is, in turn, fueled by migration and mass media. Migration is nothing new in itself: throughout history we have seen people leaving their homes, either to flee from danger or in order to pursue a better life. But while migrating used to imply being cut off from your native culture, modern mass media enables people to more easily take their culture with them and stay in touch with their homeland, via (satellite) television and internet. According to Appadurai, instead of being homogenized, globalization is more likely to result in a diversification of societies. Although there might indeed be a McDonald's in

every major city in the world, in the local food market around its corner you will find foods, ingredients and dishes from a wide variety of cultures. Moreover, he discusses how elements that were once 'foreign' to a culture can be used over time to confirm cultural identity; the (originally British) sport cricket, for instance, is now an integral part of Indian national identity. This back-and-forth between homogenization and diversification—sometimes referred to in the single term 'glocalization'—is a central theme in contemporary philosophy of culture.

As a result of the process of globalization, as well as diversification, culture is becoming the subject of sometimes intense social and political debates. Take the debate in the Netherlands, France and other European countries, about the issue of women wearing headscarves in public functions, or even the ban of the Burka from public space altogether: while some people regard the headscarf as a symbol of women's oppression, and therefore inappropriate within 'our western civilization', others stress that there are girls and women who say they wear headscarves by choice, to express their religious and cultural identity. These individuals demand respect for what makes them different from Western culture, and argue that banning headscarves is actually a form of intolerance toward different cultures and religions.

We see here that the different meanings of the concept of 'culture' that we discussed in the first section are intertwined. While one group relies on the value of emancipation and progress, and views the headscarf as 'uncivilized', the other relies on the value of a unique and authentic expression of one's identity. Sometimes, we even find these two cultural concepts within one point of view. For example, if a right-wing politician says that he finds a different culture 'backwards' in comparison to our superior western culture, then he uses both the Herderian concept of unique cultures (that is, the idea that cultures radically differ from each other) *and* the enlightenment concept (namely that cultures can be placed along one and the same yardstick of 'civilization').

Globalization of cultures leads to tensions, not only between cultures but also within them. Cultures, and the traditions and beliefs that are part of them, are constantly changing, but the pace at which this is happening can cause feelings of anxiety and alienation for some parts of the population. Some even speak of 'culture wars' taking place. Marx' class division between the bourgeoisie and the proletariat, it is sometimes argued, has been replaced by the 'winners' and 'losers' of globalization, the cosmopolitans that benefit from easy travelling and free trade, and a precarious class losing their jobs as a result of the outsourcing of production or the austerity politics following the financial crises. While philosophers once hoped that globalization would turn the world into a 'global village' (McLuhan 1964), there seems to be a countertendency towards new forms of nationalism and tribalism, the deepening of divides between (ethnic, religious,

social) groups, and an increasing hostility amongst them. Philosophers of culture try to understand these tendencies, their origins, and potential consequences.

## 6. CONCLUSION: BEYOND BORDERS, BEYOND CULTURE?

One of the problems of the aforementioned resurgence of nationalism and tribalism is that most challenges facing us today are not confined to the borders of a single nation state. Issues like climate change, pollution, nuclear power, terrorism, economic crises, and culture wars transcend the borders of nation states and intervene in the lives of people from very diverse backgrounds. At the same time, we see how groups within a nation state no longer feel automatically connected by a national identity, language or culture and obtain their information from a variety of media and information flows.

This poses a challenge for the philosophy of culture, which itself has to move beyond borders, including its own disciplinary borders. Contemporary philosophy of culture is, more than ever, intertwined with sociology and political theory, economics and even the natural sciences. Contemporary philosophers of culture write on labour, debt and neoliberalism—issues that one would rather expect from economists. Still, each of these issues has its own particular cultural dimension. Neoliberalism is not only viewed as an economic system or model, but also as a specific perspective on humankind in its relationship to its surroundings: it considers the individual as the basic unit of human experience and harbours a distrust against the social, as a potential limitation to the freedom of the individual. According to its critics, this leads (amongst other things), to a culture of individualization and competition, also outside of the market.

The scope of contemporary philosophy of culture is further broadened as a consequence of the increasing importance of the symbolic or expressive value of commodities: a sneaker or a smartphone is much more than just a shoe or means of communication. Hence, in what is sometimes called ‘cultural capitalism’, the entire economy seems to revolve around attributing meaning and significance to ourselves and to the world around us. To Raymond Williams’ remark that ‘culture is ordinary’ we can add that culture is everywhere.

At the same time, however, the limits of what human culture can or should be becomes very clear in the age of climate catastrophe, or what is sometimes referred to as the Anthropocene (of *antropos*, the Greek word for human—i.e. the era of humankind). As philosophers like Donna Haraway (2016) and Timothy Morton (2016) argue, ecological issues show us that the distinction between nature and culture is problematic to begin with, since our human culture is from the very outset intertwined with our natural surroundings. Even the word ‘culture’ is derived from agriculture—the working of the land—the start of which once

formed the basis of civilization as we know it. In that sense, the Anthropocene forms a challenge. A challenge, first of all, for culture—particularly western culture, which is based on growth, mass consumerism, and fossil fuel energy supplies that contemporary climate scientists now claim to be unsustainable and destructive—but certainly also a challenge for philosophers of culture, to once more rethink the concept of culture in its relationship to nature and to describe, understand, and imagine ways in which human beings live with each other and their natural and artificial surroundings.

## BIBLIOGRAPHY

- Appadurai, A. (1996). *Modernity at large. Cultural dimensions of globalization*. University of Minnesota Press.
- Appiah, K. & Anthony. (2018). *The lies that bind. Rethinking identity*. Profile Books.
- Arnold, M. (1869). *Culture and Anarchy. An Essay in Political and Social Criticism*.  
<https://www.gutenberg.org/cache/epub/4212/pg4212.html>
- Baudrillard, J. (1995). *Simulacra and simulation* (S. F. Glaser, Trans.). University of Michigan Press. <https://doi.org/10.3998/mpub.9904> (Original work published 1953)
- Bauman, Z. (2000). *Liquid Modernity*. Polity Press.
- Beauvoir, S. (2015). *Extracts from the second sex* (C. Borde & S. Malovany-Chevallier, Trans.). Vintage Books. (Original work published 1949)
- Benjamin, W. (2008). *The work of art in the age of its technological reproducibility and other writings on media* (M. W. Jennings, T. Y. Levin, & Doherty, Eds.; R. Livingstone, H. Eiland, Others, & E. Jephcott, Trans.). The Belknap Press of Harvard University Press. (Original work published 1936)
- Cixous, H., & Clément, C. (1986). *The newly born woman* (B. Wing, Trans.). University of Minnesota Press.  
[https://doi.org/10.4324/9780203408483\\_the\\_newly\\_born\\_woman](https://doi.org/10.4324/9780203408483_the_newly_born_woman)
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics [1989]. In *Feminist legal theory* (Vol. 140, pp. 57–80). Routledge.  
<https://doi.org/10.4324/9780429500480-5>
- Foucault, M. (1982). The subject and power. *Critical Inquiry*, 8(4), 777–795.  
<https://doi.org/10.1086/448181>
- Hall, S. (1993). Encoding, decoding. In S. During (Ed.), *The cultural studies reader. Second edition* (pp. 507–517).
- Haraway, D. (2016). *Making kin in the chthulucene*. Duke University Press.
- Hegel, G. W. F. (1991). *Elements of the philosophy of right* (A. W. Wood, Ed.; H. B. Nisbet, Trans.). Cambridge University Press. (Original work published 1812)
- Heidegger, M. (1996). *Being and Time* (J. Stambaugh, Trans.). (Original work published 1927)
- Horkheimer, M., & Adorno, T. (2002). *Dialectic of enlightenment. Philosophical fragments* (E. Jephcott, Trans.). Stanford University Press. (Original work published 1947)
- Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, 30(2), 225. <https://doi.org/10.2307/1344358>
- Marx, K. (1978). Zur kritik der politischen ökonomie. In *Marx Engels Werke* (Vol. 13, pp. 3–160). Dietz. (Original work published 1859)
- McLuhan, M. (1964). *Understanding media. The extensions of man*. Routledge.
- Morton, T. (2016). *Dark Ecology. For a logic of future coexistence*. Columbia University Press.

- Plessner, H. (2019). *Levels of organic life and the human. An introduction to philosophical anthropology* (M. Hyatt, Trans.). Fordham University Press. (Original work published 1975)
- Rosa, H. (2013). *Social Acceleration. A New Theory of Modernity* (J. Trejo-Mathys, Trans.). Penguin.
- Said, E. (2003). *Orientalism*. Penguin. (Original work published 1987)
- Virilio, P. (2006). Speed and Politics. In M. Polizzoti (Trans.), *Paul virilio* (pp. 45–60). Routledge. <https://doi.org/10.4324/9780203007631-10>
- Williams, R. (1976). *A vocabulary of culture and society*. Routledge. <https://doi.org/10.4324/9781315729367-9>
- Williams, R. (1977). Culture is ordinary. In A. Gray & J. McGuigan (Eds.), *Studying culture: An introductory reader* (pp. 5–14). Arnold.

### RECOMMENDED FURTHER READING

- Boomkens, R. (2011). *Erfenissen van de Verlichting: Basisboek cultuurfilosofie*. Boom.
- Bloois, J. de, Cauwer, S. D., & Masschelein, A. (Eds.). (2017). *50 key terms in contemporary cultural theory*. Pelckmans Pro.
- Hellemans, B. (2017). *Understanding culture*. Amsterdam University Press. <https://doi.org/10.1017/9789048530090>
- Konersmann, R. (2003). *Kulturphilosophie Zur Einführung*. Junius.

HÉLÈNE CIXOUS





## MAX HORKHEIMER AND THEODOR ADORNO

### *DIALECTIC OF ENLIGHTENMENT*

---

**D***ialectic of Enlightenment* (1947) was written during the Second World War by Max Horkheimer and Theodor W. Adorno—two German Jews living in exile in the United States. It is an extremely rich and dense (but also dark and gloomy) work, which presents a radical critique of western culture and thought from Greek antiquity up until the horrors of the Second World War. Horkheimer and Adorno argue that those horrors and atrocities should not be viewed as a sudden outburst of barbarism or a temporary straying from the path of progress, but rather, on the contrary, as an outcome of exactly this 'progress'. The Enlightenment, which had promised to emancipate humankind, had relapsed into myth and perverted into an 'instrumental rationality' which is alienated from nature and even construes human life to means-ends relationships. The essence of rational thought is what they call 'identity-thinking': subsuming phenomena that are essentially different under one common denominator. Although subsumption is necessary to make thought possible, it must always be accompanied by reflection. If that does not happen, then anything particular (or in their terms, 'non-identical') to a phenomenon falls prey to abstraction: the rabbit in a laboratory is considered as a mere exemplar of its species, rather than as a living being. According to Horkheimer and Adorno, this not only leads to philosophical and epistemological errors but also to social and political abuse; the suppression and destruction of certain groups is the result of this reflectionless identity-thinking. However, the authors do not want to give up on enlightenment; rather, they want to enlighten enlightenment itself. In other words, this thinking must be made aware of its own dark side. Only then, can the original promise of enlightenment—a truly humane society—be fulfilled.

---

Herman Philipse

**Religious Beliefs: Can  
they Be Legitimate?**

# RELIGIOUS BELIEFS: CAN THEY BE LEGITIMATE?<sup>1</sup>

## *An Introduction to Philosophy of Religion*

Herman Philipse  
[h.philipse@uu.nl](mailto:h.philipse@uu.nl)

### 1. INTRODUCTION

Is it reasonable for intelligent and educated people in our times to endorse a religious belief that a particular god exists? Some of you may nod and point to the many positive aspects of doing so, such as emotional satisfaction, helping the poor, or constructing a coherent community of believers, who often assist each other affectionately. Others might protest by mentioning more problematic features, like religiously motivated devastating deeds.

For example, on 7 January 2015 two Islamist brothers Kouachi entered the offices of the French satirical journal Charlie Hebdo in Paris, armed with guns and other weapons. Shouting “Allahu Akbar!” they killed twelve people and injured eleven. The main motive for this massacre was a religious one: to punish journalists for their jocular depictions of prophet Muhammad. Some Muslims praised the Kouachi’s as heroes of Islam, but most Islamic organizations condemned their actions as anti-Islamic. Clearly, endorsing a specific religious belief may have positive but also evil impacts.

Philosophers of religion raise and try to answer many different questions about religion(s). An important set of such questions is concerned with religious beliefs: what exactly are they about, and can holding them be considered reasonable

---

1. I am very grateful to friends and colleagues who commented critically on this chapter, in particular to Peter Hacker, Martin van Hees, and Jeroen Hopster.

or legitimate in some sense? If we want to investigate the reasonableness of religious beliefs, we have to start by making some conceptual distinctions.

First, the word 'belief' may be used in two different ways. On the one hand, when a belief is called courageous, tentative, firm, or cowardly, we refer to someone's *believing* something, that is, to an aspect of one's mental life. On the other hand, if we call a belief true, false, likely, or improbable, we focus on its propositional content, that is, on *what* is believed. Whenever we discuss the legal legitimacy of endorsing religious beliefs, or when we practise the social sciences of religion, we zoom in on the former feature, as I did in the second paragraph. The latter aspect is prominent when we pursue ontology, epistemology, or metaphysics, and wonder how probable it is that a specific religious belief be true, whether it constitutes knowledge, or what it implies with regard to the ultimate nature of reality. In order to investigate these issues, the content of the relevant belief has to be spelled out with some precision.

Secondly, criteria of reasonableness with regard to beliefs differ depending on the area at issue. For example, given the constitutional freedom of religion in Western countries, it will be legally legitimate to endorse a specific religious belief and to go through the relevant rituals. However, doing so may be unreasonable from an ontological and epistemological perspective, if it is improbable that the belief in question is true and no convincing evidence or arguments for its truth are available. Furthermore, criteria for the epistemic legitimacy of beliefs differ depending on the relevant domain or discipline to which the beliefs belong, such as mathematics, the empirical sciences, law, ethics, or metaphysics. What are these criteria with regard to religious beliefs?

In this chapter on the philosophy of religion, I focus on the content of religious creeds, and investigate their legitimacy in the epistemological sense. We will wonder whether there are any reliable primary sources of religious beliefs (section 2). Since there do not seem to be such epistemic sources, and because many alleged divine revelations are falsified empirically, religious explanations have been eliminated gradually from the sciences (section 3). Given the resulting separation between science and religion, pious philosophers have developed various non-science-based strategies for justifying their religious convictions. After providing an overview of these strategies (section 4), I examine an argumentative strategy (section 5). Finally (section 6), we shall invite readers to make up their mind with regard to religious beliefs.

## 2. EPISTEMIC SOURCES OF RELIGIOUS BELIEFS

Let us define a religious belief as the conviction that a god, angel, devil, surviving soul, or other supernatural being really exists. According to a demographic investigation of *The Global Religious Landscape* by the Pew Research Center, published in 2012, 84% of humans on Earth were religiously affiliated in 2010. Given the fact that, on average, religious believers tend to have more children than the religiously unaffiliated and to educate their children religiously, later research on *The Changing Global Religious Landscape* published in 2017 projects that in 2060 the percentage of religiously affiliated humans on Earth will be 87,5%. If we also assume that most of the religiously affiliated endorse a specific religious belief as defined above, the question as to whether and how a religious creed can be epistemologically legitimate is of worldwide importance.

In general, a belief that  $p$  is legitimate in the epistemological sense if there are convincing grounds for concluding that  $p$ 's truth is more likely than  $p$ 's falsity. Legitimacy in this sense may be a matter of degree, and which degree is considered to be sufficient will depend on the context and the topic under consideration.

In order to investigate whether a belief that  $p$  is legitimate, we should trace the sources on which it relies, which may be primary or secondary. An epistemic source is primary if it enables us to discover whether or not it is true that  $p$ . For example, our eyes, ears, and other perceptual organs are the primary sources of beliefs about our environment, such as the belief that there is a tree in front of us. Calculating is the primary source of an arithmetic belief, like the claim that  $13 \times 24 = 312$ . Epistemic sources are secondary if they rely only on results obtained by other sources, for instance, if we hear the testimony of other human beings who claim to have detected that  $p$ .

Let us now wonder which kind of epistemic source we learned to rely on when we were taught to adopt a specific religious creed during our infancy. Did we learn to investigate whether any god or other supernatural being exists by exploring primary sources of religious beliefs? Or were we taught to trust our parents and religious teacher, such as a guru, pastor, priestess, imam, or rabbi, so that we had to rely on a secondary source? In order to practice the first approach, our teachers would have had to confront us with a great diversity of religious convictions, such as the core doctrines of Catholic Trinitarianism and Tantra Hinduism, and to instruct us how one can find out whether any of these doctrines is true. Clearly, if interpreted literally, these two religious faiths cannot both be true, since they contradict each other. Whereas monotheist religions contend that there is only one god, who consists of three persons according to the Christian doctrine of

the Trinity, a popular version of Tantra Hinduism holds that there are 33 “crore” (= 10million) deities”, that is 330 million gods.

Are there any reliable primary sources or methods of investigation that enable us to discover which of these incompatible religious doctrines is true, or whether there is a true religious belief at all? Philosophers of religion such as the American celebrity Alvin Plantinga have argued that Christian believers do not need any methods of investigation, if at least God has implanted religious beliefs in their minds or hearts. If God did so, these religious beliefs would be warranted, and no further arguments or investigations would be required in order to show their epistemic legitimacy (Plantinga, 2000; 2015). However, as soon as Christians discover, for example, that Tantra Hindus endorse religious beliefs that are contradictory to theirs and hold that these beliefs are implanted in their hearts by millions of gods, this discovery is a rebutting defeater of the Christian belief.

Nowadays, all well-educated religious believers will be aware of the plurality of religions on Earth. Consequently, they cannot trust the epistemic legitimacy of their own religious convictions, unless it can be shown by reliable research and primary sources that these beliefs are more likely to be true than each of the competing ones, including universal atheism. Are there any reliable methods of religious research? If so, what are they, and what is revealed if one applies them properly?

As the 19th century American philosopher and psychologist William James stressed already in his first lecture on *The Varieties of Religious Experience* (1902), ordinary religious believers simply follow the conventional observances of their country or community, which are communicated to them by cultural traditions. In other words, they rely on secondary sources. Since the primary sources of religious beliefs allegedly give us access to something supernatural, such as the omnipotent god of Christianity or Islam, one might expect that these primary sources differ drastically from the sources of natural knowledge. Indeed, both according to James and to many recent psychological investigations, the overwhelming “original experiences” of religious originators should be classified as “abnormal psychical visitations” (Murray et al., 2012). Let me discuss briefly what caused Saul’s or Paul’s conversion to Jesus on the road to Damascus as a prominent prototype of a primary epistemic source. Without this conversion, it is unlikely that Christianity would have spread within the Roman Empire.

According to the New Testament book Acts 9:1, when Saul of Tarsus went on his way to Damascus, he was “still breathing threats and murder against the disciples of the Lord”. How should we explain that Saul suddenly converted from a Jewish persecutor of the Jesus-sect to Paul, its main propagator during the first century? Acts 9:3-9 mentions four striking features of Saul’s conversion-experience on the road to Damascus: (1) “suddenly a light from heaven flashed

about him”; (2) “he fell to the ground”; (3) “heard a voice” that is attributed to Jesus; after which (4) “for three days he was without sight, and neither ate nor drank”. Three further features are relevant as well. Although in Acts 9:7 it is said that the “men who were travelling with him stood speechless, hearing the voice but seeing no one”, Acts 22:9 quotes Saul himself, saying that (5) “those who were with me saw the light but did not hear the voice”. In Acts 26:13-14, it is added that (6) the light from heaven was “brighter than the sun”, and that (7) all those who journeyed with Saul also fell to the ground.

If we assume that the descriptions of Saul’s conversion experience in Acts are historically reliable to some extent, we should wonder what is the best explanation of features 1, 2, 4, 5, 6, and 7, neglecting 3 for a moment, because according to 5 Saul’s companions did not hear a voice. There is a long history of psychiatric explanations of what happened to Saul, such as that he suffered from an epileptic attack or from paranoid schizophrenia, but these explanations do not account for features 1, 5, 6, and 7. A somewhat superior secular explanation of what happened to Saul has been proposed only recently, and it relies on a careful comparison between the passages in Acts and eyewitness accounts of “the explosive entry of an asteroid fragment over Chelyabinsk in 2013” (Hartmann, 2015, p. 368). What happened to Saul and his companions may have been caused by the nearby impact of an asteroid or meteorite as well.

For example, features 2 and 7 can be explained by the shockwave resulting from an asteroid exploding when it descends within the Earth’s atmosphere, and aspects 1, 4, 5, and 6 are accounted for by the intense ultraviolet radiation caused by such a fireball event. If this is indeed what happened, Saul’s ignorance of the astrophysical knowledge available today explains that he interpreted this perplexing experience in a religious way. Since he perceived the fireball event as the resurrected Jesus appearing to him (3), it is plausible to assume that Saul looked longer into the intense ultraviolet radiation than his companions, whereby he got photokeratitis (feature 4). The content of Saul’s interpretation, to the effect that (3) he heard the voice of Jesus resurrected, and Saul’s radical conversion resulting from his experience, have been explained convincingly by Friedrich Nietzsche in *Morgenröte* (1881, §68), although Nietzsche knew nothing about asteroids, and endorsed a purely psychiatric diagnosis.

The secular (asteroid) account of what happened to Saul on the road to Damascus overrules the religious explanation, which many Christians still endorse. Quite probably, Saul’s conversion was not based upon any trustworthy primary epistemic source of religious beliefs. Rather, it resulted from Saul’s not knowing the real causes of his overwhelming ordeal.

### 3. SCIENCE AND RELIGION

As the St. Paul-example illustrates, during our long (pre-) history, human ignorance concerning the causes of striking events such as thunderbolts, epidemics, or floods triggered their attribution to divine beings. Many other motives gave and still give rise to supernatural beliefs as well, like longing for a life after death. As David Hume argued in *The Natural History of Religion*, it is “[n]o wonder [...] that mankind, being placed in such an absolute ignorance of causes and being at the same time so anxious concerning their future fortunes, should immediately acknowledge a dependence on invisible powers, possess of sentiment and intelligence” (Hume, 1757, p. 34). Today, cognitive scientists have proposed supplementary explanations of religious beliefs, such as the hypothesis that our brains contain a Hypersensitive Agent Detection Device. This mental mechanism would have enhanced evolutionary fitness, although it also gave rise to supernatural illusions.

If we define ‘science’ in the broadest sense as the human pursuit of knowledge by means of well-calibrated methods of research, we will understand why from the scientific revolution in 16th-17th century Europe onwards, religious beliefs have been eliminated gradually from science-informed world views. It became ever more obvious that no religious belief could be justified from an epistemological point of view, so that we should wonder why it might be legitimate to endorse such a belief at all. Let me just mention the most important reasons why religious beliefs were eliminated from well-informed world views.

First, numerous religious explanations of empirical phenomena have been overruled by scientific ones because the latter were epistemically superior in many respects, such as testability, falsifiability, and fecundity for further research. Whereas Newton still thought that cosmological features like the fact that all planets move “in Orbs concentrick” around the Sun in the same direction, could be explained only by assuming that God decided to create them like this (Newton, 1730, Query 31, p. 402), both Immanuel Kant and Pierre Simon de Laplace provided purely physical explanations of these features, such as their nebular hypothesis. Similarly, although William Paley still argued in his *Natural Theology* (1802) that the functional complexity of living beings and their parts (e.g. the eye) could be explained only by the Christian doctrine that God had created the first instances of each species, Darwin refuted this creed of special creation in *The Origin of Species* (1859) by showing extensively the empirical superiority of his evolutionary hypothesis.

A second reason for eliminating religious beliefs from our scientific view of the world is that many of the factual claims contained in collections of sacred scriptures such as the Bible or the Koran have been shown to be false.



For example, whereas according to the Biblical book *Genesis* I, God created the universe primarily in order to house humanity, present-day cosmological estimates of the actual size of the universe as it has expanded since the Big Bang demonstrate the absurdity of this homo-centric prejudice. Similarly, the many Bible-inspired calculations of the age of the universe, such as Archbishop Ussher's computation on the basis of biblical genealogies to the effect that God started to create the world in October 4004 BC, have been refuted by numerous results of scientific research. For instance, recently developed radiometric dating techniques have revealed that planet Earth is approximately 4.45 billion years old, and evolutionary biologists estimate that life began on Earth between 4.2 and 3.5 billion years ago. Given the falsity of many factual claims contained in allegedly holy books, it is highly unlikely that they were revealed to humanity by an omniscient deity.

Thirdly, the more scientists and philosophers reflected on the epistemic merits of methods for detecting truths, the sharper they realised that no religious belief is obtained or justified by using any trustworthy method or reliable primary source. This insight is supported as well by the fact that mutually incompatible religious beliefs are endorsed by humans with reference to the same types of alleged primary sources, such as holy books or divine revelations during dreams. From these two results it follows, fourthly, that given the plurality of mutually incompatible religious beliefs, there are no primary epistemic sources or methods available that enable us to discover that any one of these convictions is true, or is more likely to be true than each of its religious rivals. As a consequence, religious hypotheses should be, and have been, eliminated completely from the scientific enterprise.

One might object that scientific progress not only eliminates religious accounts of many empirical facts, since these accounts clearly resulted from ignorance, as Hume argued. Scientific progress also reveals ever more things the causes or origins of which we do not know. It has been suggested that some of these things, such as the postulated Big Bang, should be explained by a religious hypothesis. However, both Christians like Henry Drummond, and many atheists have condemned this proposal as a "god-of-the-gaps" approach. For religious believers the strategy is too risky, because one cannot exclude the possibility of scientific progress eliminating the relevant epistemic gap and its religious interpretation. Atheists will maintain that no religious explanation of any fact will be acceptable anyway, since there are no reliable methods of religious discovery. Many experts have concluded that from an epistemological point of view, science and religion should be separated sharply. For instance, Stephen Jay Gould argued in his 1997 essay "Nonoverlapping Magisteria" that science and religion(s) are concerned with different domains that do not overlap: the domain of facts (sciences) and the domain of values (religions). Will this strategy of a

radical separation between religion(s) and science be epistemically beneficial for religious believers? What do you think?

#### **4. LEGITIMISING RELIGIOUS BELIEFS. A DECISION TREE FOR THE FAITHFUL**

**S**tephen Jay Gould's separation strategy confronts religious believers with a difficult dilemma. Either (a) they agree that their religious belief should not be interpreted as a factual claim to truth. Consequently, they should not endorse any more the conviction that a particular god exists in fact. However, this abstention will undermine many of the justifications they adduce in favour of their values. Or (b) they should argue, against Gould, that what has justified the elimination of religious beliefs from the scientific enterprise is not the epistemic unreliability of their primary sources. In order to substantiate this second option, they have to show what these sources are, and how using them justifies the conviction that in fact a specific god exists.

Followers of Søren Kierkegaard will (c) try to escape between the horns of this dilemma by holding that given the many apparent paradoxes of a Christian (or other religious) creed, which transcend our rational capacities, one can (and should) endorse the factual truth of this (or any other particular) religious doctrine by suspending reason completely, and believing it by virtue of the absurd. Only if one does so in all practices of life, can one become one's true self. In other words, our religious belief should radically overrule reason.

One might model the choice between options a, b, and c as the first node (I) of a decision tree for religious believers. Each of these options may be pursued in different manners, so that they lead to new nodes at level (II). For example, if one endorses (a), and thus embraces a so-called non-cognitivist interpretation of religious beliefs (which arguably do not aim at knowledge), one will re-interpret the conventional utterances of one's creed, such as 'God is omnipotent and created the universe', or 'Poseidon will save me from drowning', either as (1) mere expressions of emotions, or (2) as implicit moral claims, (3) as affirmations of belonging to a specific community, or (4) as having these and many other functions that are distinct from making a propositional claim to truth. These are so-called 'revisionary accounts' of religious language, since most religious believers intend to make factual statements when they assert that their god(s) exist(s) and did or will do something.

According to 'non-revisionary accounts' of religious language, existence claims with regard to specific gods such as Yahweh, God, Allah, Zeus, or Zhenwu typically aim at stating a matter of fact. Both options (b) and (c) endorse the so-called face value theory of religious language. They take religious utterances at

face value, and interpret them as closely as possible to what they seem to mean at first sight. Since option (c) denies the relevance of religious epistemology, I shall focus here on option (b), which is a node for the following dilemma at level II. If (b) a religious statement is a factual claim to truth, either (5) endorsing it can be epistemically legitimate only if based upon relevant types of evidence, or (6) no evidence is needed for its epistemic legitimacy. As Alvin Plantinga (2000, 2015) argued, (6) is correct for Christians if God exists, since, if so, probably He implanted properly Christian beliefs into the minds of Christian believers. Hence, these beliefs will be warranted even if no evidence whatsoever supports them.

If we think that evidence *is* needed in order to endorse legitimately the factual claim to truth that God exists (5), we arrive at the node for the next dilemma, at level III, which is concerned with the types of evidence that are relevant for legitimising religious beliefs like this. According to the first horn of this dilemma, option (7), the evidence for a specific religious belief will become available only to those who engage receptively within the relevant religious community for quite some time (e.g. Cottingham, 2014). As a consequence, Cottingham concludes, one cannot reach “a final verdict on the relevant truth-claims from outside the forms of practical and affective engagement through which alone genuine understanding flourishes” (2014, p. 171).

Both options (6) and (7) run into trouble if confronted by the problem of the plurality of religions. For instance, the Christian belief that there is only one god, called God, who is omnipotent and omniscient, is contradicted by polytheistic religious beliefs, such as the Tantra-Hindu conviction that there are 330 million gods, so that not each of these beliefs can be true. If representatives of both religions choose option (6), according to which no evidence is needed for endorsing legitimately a religious belief, they will have to conclude that the truth of the belief they endorse is very unlikely, given the large quantity of mutually incompatible religious beliefs. The same holds for option (7), because adherents of many different and mutually incompatible religious convictions will say that they have discerned the truth of their belief only after a long-lasting involvement within the relevant community.

On these and other grounds, many religious apologists will prefer the second horn of the dilemma at level III, according to which (8) the evidence for a specific religious belief (and against the competing religious beliefs) is accessible to non-believers as well, at least to some extent. In other words, some primary sources of religious beliefs must be publicly available, and it should be possible to justify a specific religious belief by using these sources and reliable methods of research.

However, given the elimination of religious beliefs from the scientific enterprise, illustrated in section 2 above, option (8) is quite challenging. Have religious beliefs

not been excluded from the sciences because no reliable methods of research or trustworthy primary epistemic sources could support any of them? Will not each piece of empirical evidence adduced to the credit of a specific religious belief run the risk of merely advocating a god-of-the-gaps, that is, postulating a religious explanation for things that the sciences have not yet explained, but might explain in the future? Since option (8) is challenging, it may be interesting to devote the next section to one of its most prominent proponents.

## 5. A SOPHISTICATED CUMULATIVE CASE STRATEGY OF ARGUMENTS FOR THEISM

One of the most sophisticated attempts to elaborate option (8) is to be found in Richard Swinburne's book *The Existence of God* (1979, 2004). Using Bayesian probability theory, Swinburne intends to show that given the total empirical evidence *pro* and *contra*, "theism is more probable than not" (2004, p. 342). He defines 'theism' as the thesis that God exists, while the word 'God' is introduced as the name of a person picked out by the following description: "a person without a body (i.e. a spirit) who necessarily is eternal, perfectly free, omnipotent, omniscient, perfectly good, and the creator of all things" (2004, p. 7).

How does Swinburne avoid the god-of-the-gaps risk of option (8) that the empirical evidence adduced in favour of theism will be explained later by superior scientific theories, so that theism will be superseded? Whereas he claims on the one hand that theism purports to explain "all our empirical data", he argues on the other hand that the empirical evidence for theism is either "too odd" or "too big" for science to explain (2004, pp. 66, 93, 74-5). Furthermore, although he states that theism is a "large-scale theory of the universe", which closely resembles "large-scale scientific theories" (2004, p. 3), he also holds that there is no risk of an empirical refutation of theism in the future, since theism "does not yield predictions such that we can know only tomorrow, and not today, whether they succeed" (2004, p. 70).

In order to use Bayes' theorem for calculating the probability that theism is true given all available evidence, many conditions have to be met. Let me mention three of them only. First, theism should be defined meaningfully, and it is debatable whether Swinburne's traditional definition of 'God' succeeds in this respect. Since we give meaning to the word 'person' by applying it to human beings who are corporeal, it may be argued that one annuls the meaning of 'person' by stating that God is a person without a body. Secondly, Bayes' theorem can be applied to theism only if specific empirical facts are more (or less) likely if God exists than if God does not exist. In other words, the hypothesis of theism should have some predictive power if it makes sense at all, and in order to grasp

this predictive power we humans should be able to know God's intentions, at least to some extent. Thirdly, in order to show that God's existence is more probable than not, one should attribute a prior (initial) probability to theism which is not too low. Since there are infinitely many different monotheistic and polytheistic doctrines, which contradict each other in many respects, one might argue that the prior probability that one of them is true approaches zero, because there are no convincing grounds for attributing different prior probabilities to these proposals. If so, the prospects of producing convincing inductive arguments for this specific version of theism are negligible.

There is no space here for explaining how Swinburne or other religious apologists have attempted to resolve these problems (cf. Philipse, 2012, 2014, chapters 7-11). Let me just mention some main types of empirical arguments put forward in support of the theistic doctrine as defined. According to (A) cosmological arguments, the existence of the universe confirms the hypothesis of theism, since allegedly it is more likely that our universe exists if theism is true than if theism is false. In arguments (B) from temporal order it is claimed that probably a Godless universe would be quite chaotic, so that the fact that our universe is governed by laws of nature corroborates theism as well. So-called fine-tuning arguments (C) start from the premise that life in general and the human species in particular can evolve only in a universe the natural laws and initial conditions of which are quite specific, compared to the large set of conceivable laws and initial conditions. Then it is argued that these facts of fine-tuning are much more probable if God exists than if God does not exist, since, so it is assumed, God intended to create a world in which humans could evolve.

In order to demonstrate by means of Bayes' theorem that theism is more probable than not, one should not only attribute a plausible prior probability to this hypothesis, but also argue that the total evidence in favour of theism overrules sufficiently the evidence against. According to theists such as Swinburne (2004), many other facts apart from (A-C) confirm theism empirically, such as (D) the overall beauty of nature, (E) the fact that conscious beings exist, (F) the complex brain-mind connections in humans, (G) our moral awareness, (H) the need for humans to make significant moral choices, and (I) the presence of religious experiences. Furthermore, theists will argue that facts (A-I) overrule other facts that seem to disconfirm theism, such as the existence of (J) moral evil and the many kinds of (K) natural evils.

Critics may reply that theism is not at all confirmed by facts (A-I), whereas there are many more empirical arguments against theism than its proponents take into account. As scientific progress in cosmology reveals, for example, life is extremely rare in our universe (L). God would never have created such a cosmos, if He were perfectly good in Swinburne's sense. Furthermore (M), since monotheism arose only rather late in human (pre-) history, if God existed he

would have been hiding himself from humans during many millennia, and does not reveal Himself to each human individual today. Because allegedly God is a good father for all human beings, he would not conceal himself for anyone of us, so that facts of divine hiddenness show that God does not exist (Schellenberg 2006, 2015). Furthermore, as neuro-scientific research has shown ever more convincingly (N), mental phenomena can exist only on the basis of specific neural substrata, so that it is highly unlikely that God as defined by theism exists at all.

## 5. MAKING UP YOUR MIND

Questions as to whether a specific divinity exists are not only interesting from an epistemological point of view. It can be of great practical importance whether one endorses or refuses to sustain a particular religious belief. As Daniel Dennett (2006, p. 15) stresses, “for many people, probably a majority of the people on Earth, nothing matters more than religion”. Hence he “can think of no more important topic to investigate” (p. 7). Whereas Dennett focuses on the evolutionary, social, economic, and psychological functions of religious convictions, the core question of the present chapter is an epistemological one: can it still be reasonable for intelligent and educated people in our times to endorse a religious belief that a particular god exists?

Since there is no consensus between experts on this issue, each of you will have to make up your mind yourself. This chapter aimed at providing an overview of possible options in this respect, and motivating readers to make up their mind with intellectual integrity.

## BIBLIOGRAPHY

- Cottingham, J. (2014). *Philosophy of religion. Towards a more humane approach*. Cambridge University Press.
- Darwin, C. (1996). *The origin of species. Edited with an introduction and notes by Gillian Beer*. Oxford University Press. (Original work published 1859)
- Dennett, D. (2006). *Breaking the spell. Religion as a natural phenomenon* (L. A. Lane, Ed.). Penguin Books.
- Gould, S.J. (1997). Nonoverlapping magisteria. *Natural History*, 106, 16–22.
- Hartmann, W. (2015). Chelyabinsk, Zond IV, and a possible first-century fireball of historical importance. *Meteoritics & Planetary Science*, 50(3), 368–381.  
<https://doi.org/10.1111/maps.12428>
- Hume, D. (1976). Dialogues concerning natural religion. In A. W. Colver & J. V. Price (Eds.), *David hume: The natural history of religion and dialogues concerning natural religion*. Oxford University Press. <https://doi.org/10.1093/oseo/instance.00044084> (Original work published 1779)
- Hume, D. (1976). The Natural History of Religion. In A. W. Colver & J. V. Price (Eds.), *David hume: The natural history of religion and dialogues concerning natural religion*. Oxford University Press. <https://doi.org/10.1093/oseo/instance.00044084> (Original work published 1757)
- James, W. (1977). *The varieties of religious experience (the Gifford lectures delivered at Edinburgh 1901-2)*. Collins. (Original work published 1902)
- Murray, E., Cunningham, M., & Price, B. (2012). The role of psychotic disorders in religious history considered. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 24(4), 410–426. <https://doi.org/10.1176/appi.neuropsych.11090214>
- Newton, I. (1979). *Opticks or a treatise of the reflections, refractions, inflections & colours of light*. Dover Publications. (Original work published 1730)
- Nietzsche, F. (1881). *Morgenröte. Gedanken über die moralischen vorurteile*. Alfred Kröner Verlag.
- P Paley, W. (2006). *Natural theology: Or, Evidences of the existence and attributes of the deity, collected from the appearances of nature* (M. D. Eddy & D. Knight, Eds.; 2nd ed.). Oxford University Press. <https://doi.org/10.1037/11747-000> (Original work published 1802)
- Philipse, H. (2014). *God in the age of science? A critique of religious reason*. Oxford University Press. (Original work published 2012)
- Plantinga, A. (2000). *Warranted Christian belief*. Oxford University Press.  
<https://doi.org/10.1093/0195131932.001.0001>
- Plantinga, A. (2015). *Knowledge and Christian belief*. William B. Eerdmans Publishing Company.
- Schellenberg, J. (2006). *Divine hiddenness and human reason. With a new preface*. Cornell University Press.
- Schellenberg, J. L. (2015). *The hiddenness argument: Philosophy's new challenge to belief in God* (First edition). Oxford University Press.
- Swinburne, R. (2004). *The existence of God* (2nd ed.). Oxford University Press. (Original work published 1979)

BLAISE PASCAL





## SOME CLASSICS ON THE PHILOSOPHY OF RELIGION

---

If defined broadly, the Philosophy of Religion involves issues in all the main areas of philosophy, such as metaphysics, ethics, philosophy of art, of language, science, and so on. Since it is more instructive to discuss one of these areas only with some precision instead of providing an overview, I focused on religious epistemology in this article. Let me refer readers interested in an encompassing survey of the field to sections on the Philosophy of Religion in encyclopedias available on internet, such as the *Stanford Encyclopedia of Philosophy*.

To those of you who want to read a well-written introduction, I would recommend Robin Le Poidevin, *Arguing for Atheism. An Introduction to the Philosophy of Religion* (Routledge, 1996). In our times, philosophers who are believing Christians have defended their faiths in many different ways. One of the most able advocates of theism is Richard Swinburne, whose book *The Existence of God* (second edition, 2004) I recommend to those of you who are interested in rational arguments in favour of a religious claim to truth. I evaluated Swinburne's arguments in my (2012, 2014) book *God in the Age of Science? A Critique of Religious Reason*. A very different Christian approach to religious beliefs has been defended by John Cottingham in his *Philosophy of Religion. Towards a More Humane Approach* (Cambridge University Press, 2014). Inspired by Blaise Pascal, he pleads for "an epistemology of involvement" if one wants to understand God's personal call to allegiance, which is directed "not to the analytic mind but to the heart".

Let me end my list of recommended books on philosophy of religion with a classic written by John Leslie Mackie: *The Miracle of Theism. Arguments for and against the existence of God* (Clarendon Press, Oxford, 1982). According to Mackie, "the question whether there is or is not a god can and should be discussed rationally and reasonably", whereas this question is "too important for us to take sides about it casually or arbitrarily".

---

---

# THINKING ABOUT HUMANS

---

Monica Meijsing

# **Cartesian Conundrums**

# CARTESIAN CONUNDRUMS

## *An Introduction to Philosophy of Mind*<sup>1</sup>

Monica Meijsing  
m.meijsing@tilburguniversity.edu

### 1. INTRODUCTION

**W**hat happens when a loved one dies? A moment ago, she was still warm and alive, and now something lies dead. Surely, they cannot be the same: the person you loved, and the cold body? Where did she go?

These questions are as old as humanity, and have been answered by most religions and world views in terms of a dualism of mortal bodies and immortal souls. Such dualism comes quite naturally to people, even in present-day, secular societies.<sup>2</sup> For centuries, they were accepted by lay people and scholars alike. But the Scientific Revolution of the 16<sup>th</sup> and 17<sup>th</sup> century – often called “the mechanisation of the world view” – gave rise to a different dualism, and it was René Descartes (1596-1650) who devised it.

Where before the Scientific Revolution the whole universe was seen as a kind of living organism, and organisms as being capable of pleasure and pain, afterwards the whole universe, including the bodies of living organisms, was considered to be a non-living mechanism. But although the mechanical body

---

1. I would like to thank Martin van Hees, Lodi Nauta, Bart Engelen and Maureen Sie for their very helpful feedback.

2. Bloom (2004) argues that people are natural dualists.

is not capable of pleasure and pain, *something* surely is. And that *something* is precisely the being that thinks about what it is. Descartes therefore divided the universe in two radically different kinds of “stuff”: the physical “stuff” (*res extensa*) which consisted of the whole physical universe with everything in it, including human bodies, and the mental “stuff” (*res cogitans*), like our minds, which was capable of both thought and feeling. Unlike the “outside world” of the *res extensa*, the *res cogitans* was immediately known, a kind of “inner world”.<sup>3</sup>

With this radical sundering of the universe, Descartes set the agenda for all philosophy of mind. The first question, immediately recognised in Descartes’ own time, was that of interaction. How could the mental, thinking and feeling, ever *influence* physical matter? This is referred to as the interaction problem. During the 19<sup>th</sup> and 20<sup>th</sup> centuries, what is called physicalist monism took over as world view: everything that exists is physical. That seemingly solved the interaction problem. But the problems for philosophy of mind did not go away; they only changed form. The question now was: how can thought and feeling *be* something physical? That topic still sets the agenda for present day philosophy of mind: what is the nature of thought and of feeling? Thought is first and foremost characterised by a kind of *aboutness*: whenever we are thinking we are thinking *about* something. Our thoughts have *content*. And feeling is characterised by *consciousness*: feelings that are not felt simply are not feelings. Thus, one might say that the two main topics of philosophy of mind are content and consciousness.

In the following, we are going to look at three successive attempts to deal with these questions. We will, incidentally, see that these attempts are, just as in Descartes’ time, heavily influenced by developments in the sciences. And we will also see that the question of content and that of consciousness are seldom tackled together. Often, philosophers grapple for a while with the question of content and then, without any definitive answer being given, shift their focus towards the problem of consciousness, and back again.

## 2. FIRST ATTEMPT: BEHAVIOURISM

**D**uring the 18<sup>th</sup> and 19<sup>th</sup> centuries, the science rooted in the Scientific Revolution, gathered momentum and started to yield spectacular results that changed western society. Its prestige grew apace. In order to share in that prestige, psychology began to change its self-conception: instead of a science of the soul – the literal translation of “psychology” – it became the science of behaviour. This movement started in the beginning of the 20<sup>th</sup>

---

3. See text.

century in the US, where John Watson coined the term “behaviourism” – “a purely objective, experimental branch of natural science which needs introspection as little as do the sciences of chemistry and physics.” (Watson, 1913, p. 158).

Behaviourism meant that psychology was no longer viewed as the science of the inner soul or mind, which was once thought to be immediately known by introspection, but as the study of publicly accessible behaviour of animals and human beings only. It was thus a rule of methodology not to talk about this inner world. Psychology should try to establish lawful connections between stimuli and behavioural responses, both described in the objective terms of the natural sciences. All reference to sensations and thoughts was to be avoided. A test animal wasn't feeling hungry, it had been deprived of food for such and such a time. The subject of psychology, human or non-human, was considered to be a kind of black box: you gave it stimuli as input, and you measured the behavioural output. All reference to what went on inside the black box, to inner processes of thinking or sensation, was considered to be nothing but so-called intervening variables that were only postulated to smooth out the stimulus-response relation.

This refusal to even address the occurrence of conscious episodes in their subjects, earned the behaviourist psychologists the accusation that they were “feigning anaesthesia”: isn't it patently untrue that there are no conscious feelings? The behaviourist Skinner tried to see thought as a kind of behaviour too: it was subliminal speech, and he suspected – but never managed to establish – that thought was accompanied by tiny movements of the speech apparatus.

In philosophy, behaviourism took another form. Gilbert Ryle directly addressed the Cartesian idea of a non-material mind somehow residing in a mechanical body; the “ghost in the machine”, as he named it (Ryle, 1949). According to him, there could not possibly be an inner mental world, where the movements of the physical body were planned and devised, and subsequently caused. Ryle is often seen as the originator of *logical behaviourism*, the view that there is no need for inner causes of outer behaviour, because behaviour is all there is.<sup>4</sup> Unlike methodological behaviourism in psychology, logical behaviourism was not interested in describing behaviour in physicalist terms. All it claimed was that everything that Cartesian dualism deemed to be mental and internal, was in fact an aspect of public bodily behaviour. Saying that John thinks it will rain, means nothing more than that John is disposed to close the windows and take an umbrella when going out. Saying he is in pain, means he is disposed to show so-called pain behaviour: moaning or cursing maybe, taking an aspirin, avoiding stress to a certain part of his body. Neither thought nor feeling are ever the inner causes of outer behaviour; they are always aspects of public behaviour, or else *dispositions* to behave in a certain way.

---

4. Though Ryle himself didn't like to belong to any kind of “ism”. See e.g. Park 1994, Vrijen 2007.

The problem with logical behaviourism was that it is simply counterintuitive that there are no inner causes of outer behaviour. And if the mental is an aspect of behaviour, what about the Super-Spartan, who is in pain but never behaves like it? Or the Super-Actor who can behave angrily, without feeling any anger?

Behaviourism tried to solve the problem of interaction between mind and body by simply not mentioning, or even denying, the existence of mind, of an inner world, of inner causes for outward behaviour. Yet the theory fits poorly with our own experience of what it is to be a human being. And although behaviourism of some kind ruled supreme for some 50 years in psychology, it was finally abandoned in favour of a kind of psychology and philosophy of mind that paid more attention to what traditionally belonged to the Cartesian mind.

### 3. MEANWHILE IN SCIENCE 1: THE RISE OF COMPUTERS AND ARTIFICIAL INTELLIGENCE

**W**ill saw the beginning of a hugely influential development: that of computers that could perform all kinds of intelligence-demanding tasks. Already in 1950, Alan Turing, the pioneer of computer science, asked whether computers could *think*. Though he proposed to test this in a rather behaviourist way, by trying to see whether a computer could *behave* indistinguishably from a human being, the existence of intelligent machines gave rise to all kinds of speculations regarding Cartesian dualism. Computers are fully physical devices, without mind or soul. Yet they can perform all kinds of tasks that – for human beings – require intelligence: playing chess, solving problems, recognising patterns. While not having a mind or soul, computers do have rich inner (not outwardly visible) states and processes, and these inner states and processes do cause their outward behaviour. So, maybe the Cartesian inner world of the mental could be salvaged *without* a Cartesian dualism. One could perhaps be a *physicalist* (materialist) while still talking about *mental* states and processes. This would of course change the meaning of the concept of the mental considerably. “Mental” no longer denoted a substance distinct from the physical; it only denoted the characteristics of “inner”, “content-bearing” and “influencing (outward) physical matter” that Descartes had imputed to the *res cogitans*.

Computers were sometimes compared to the brain: both were seen as physical systems consisting of units (bits or neurons) that can be on or off, active or inactive, and whose inner processes cause the behaviour of the whole system. Others considered computers as the physical substrate, whereas the program was seen as something mental, a meaningful part distinct from the physical system that at the same time “runs” on it and controls it. Computers, especially in the latter view, seemed to license mentalistic talk, both in psychology and

philosophy. From a philosophical perspective, this was a good thing: it allowed talk about physical (hardware) and mental (program) things without accepting some Cartesian dualism, with its seemingly unsolvable interaction problem.

#### 4. SECOND ATTEMPT: BRAIN-MIND IDENTITY THEORY AND FUNCTIONALISM

In the late nineteen-fifties, a new position was developed in the philosophy of mind: the mind-brain *identity theory*. Identity theorists said that, contrary to logical behaviourism, there are inner causes for outward behaviour: these inner causes are brain states and processes. They also claimed that my headache is not just a (side) effect of a certain brain state, it is in fact *identical* with that brain state, just as the thought "I'd better get some aspirin" is identical with another brain state. As a result of these claims, a description of human behaviour can be provided in terms of a completely physical causal chain from, say, the intake of alcohol, a number of brain processes and states, to the intake of aspirin. Behaviour is *caused* by brain states and processes, but at the same time it is just as literally *caused* by thoughts. According to those who endorsed the identity theory, this identity of mental states and processes (experiences of headaches and thoughts of aspirin) and brain states and processes was an empirical discovery, just as it was empirically discovered that water is identical with H<sub>2</sub>O. (Place, 1956).

This last claim, however, is not as straightforward as it looks. Empirical studies can never *prove* this identity; all brain science can ever show is that there are *correlations* between neurophysiological states and processes and mental states and processes – and new techniques of neuro-imaging bring ever more of these correlations to the light. That criticism notwithstanding, it is easy to see the attraction of this theory. It would be a very elegant solution to the Cartesian problem of interaction if the whole notion of *two kinds* of states and processes that interact with one another, mental and neurophysiological, could be replaced with just *one kind* of states and processes. According to identity theory, it is the brain that does all the causal work. When we think "I'd better take some aspirin", it is true that thought causes us to take an aspirin, for the thought itself *is also* a brain state. In this way we can stick to talking about thoughts, desires, and other mental states as well as about causal relations between those states and our behaviour, as long as we realise that this is just a different way of talking about physical states.

Identity theory was seen to be supported by the existence of computers. Here were completely physical systems that could nevertheless perform actions that seemed to require intelligence. Descartes' claim, that at least human beings could not be wholly mechanical systems, because a machine could never do



what human beings easily can – namely, answering questions –, was suddenly disproven. Computers could do just that. The first computers from the fifties of the 20<sup>th</sup> century were popularly called giant brains (and these early computers were gigantic indeed). This comparison between brain and computer was based on the following: the brain consists of a network of brain cells, called neurons, which are connected with one another. These neurons can produce little bursts of electricity in an all-or-nothing fashion. Via the connections between neurons, these bursts are transmitted to neighbouring neurons, which can likewise “fire”. The combination of billions of neurons can eventually exhibit the mental capacities that human beings have: reading, translating, solving problems, playing games. The same goes for computers. Their basic element is the digital bit: an element that can be a 0 or a 1. It is the combination of these basic elements that gives rise to all the complex, intelligence-demanding things that computers can do. And indeed, the first kind of computers that were used for Artificial Intelligence were minimal networks of binary elements, called *perceptrons*, that were supposed to be self-organising, and could learn. These experiments with perceptrons were abandoned because of their limited success (Minsky & Papert, 1969). Other kinds of computers, with a so-called von Neumann architecture, were much more successful in AI. These had a central processor, a memory to store data and instructions, and input and output devices. Such a computer works because it is *programmed*, but the very same program can run on physically different machines.

Parallel with the dismissal of perceptrons as good models for AI, there was philosophical criticism of the identity theory. *The first kind of criticism* thought the identity theory was too restricting. If my thought of food is identical with brain process a, with characteristics F, G and H, then every thought of food has to be identical with such a brain process; just as the identity of this particular bit of water with H<sub>2</sub>O implies that every sample of water has to be identical with H<sub>2</sub>O. This has to do with what is known as Leibniz’ law: if a is identical with b, then a and b must have *all* of their characteristics in common; this is just what ‘identical’ means. But the brain of a cat or dog is different from mine, and Martians or computers have no brain at all. Does this imply that they cannot think of food? Even other human beings do not have exactly the same kind of brain that I have. Am I the only one capable of thinking of food? That seems very implausible. And finally, my own brain does not stay the same: the very brain process that I had yesterday, thinking about food, is not there anymore today: neurons change or die continually.

Identity theory was thus seen as too anthropocentric, allowing only *human* brain states to be identical with mental states. Maybe my thought of food today is identical with brain process a, while my cat’s thought of food is identical with *his* brain state b, and maybe a computer can think of food as well, and then

its thought is identical with *its* internal state c.<sup>5</sup> If one wants to do psychology, it wouldn't do to claim that no two creatures can ever have the same thought, for how would one then be able to say that in general a thought of food leads to a certain type of behaviour? But maybe all these instances of thoughts of food, though identical as thoughts of food, do not have to be identical as to their *physical* characteristics. If that can be the case, one can still be a physicalist and avoid dualism: a thought is always identical with *some* physical state. It's just that the *same* thought can, in different creatures, be identical to *different* physical states. This tallies nicely with the fact that the same computer program can run on physically different machines. In computer science this means that programs are, what is called, multiply realisable. Likewise, it is claimed that mental states are multiply realisable. The acceptance of these claims led to a new position called functionalism.

Functionalism claims that a mental state, such as the thought of food, is not defined by the physical state it happens to be identical with, but with the *functional role* it has in the causal chain. Every physical state that plays a role in the causal chain between food deprivation, feelings of hunger, and food-directed behaviour, *is* in fact a thought of food, whether this physical state is human, feline or canine. Human beings, cats and dogs are all different physical systems, but they can all think of food – and so can computers or Martians, if their internal states play the same role. Just as a mousetrap is defined by what it *does*, and not by precisely what physical system it is, so a mental state is defined by what it *does*, and not by what physical (brain) state it happens to be identical with on a certain occasion. Instead of saying that mental states are identical to brain states – but the same mental state can be identical to *different* brain states on different occasions – they prefer to say that mental states are *realised* by brain states (or computer states).

According to functionalism, the identity theory had seen the man-machine comparison in the wrong way: it was not about brains as computers, but about what human beings and computers can *do*. One could study mental processes without bothering too much about the physical realisation of those processes, because mental processes and mental states are multiply realisable. The physical constitution of the brain was considered just as irrelevant as the precise physical constitution of a machine. What was important was how they work on a more abstract level. What both computer and human beings (and other animals) do is performing computations on internal states, in other words, symbol manipulation. One could study the mental directly, without bothering about the physical realisation (Levin, 2016).

---

5. One could even claim that no two thoughts of food are ever exactly identical, as thoughts are thoroughly context-dependent. A full discussion of this topic would lead too far afield, but see e.g. Fodor and Lepore (1992).

*The second kind of criticism* on identity theory went into the opposite direction. If there is an identity of two things, there actually are not *two* things but *one*. If Shakespeare is in fact identical with Francis Bacon, there are not two men, one a playwright and one a statesman, but one man who is both. So, if mental states are identical with brain states, all that really exist are brain states. What we have learnt is not that mental states are identical to brain states, but that what we thought were mental states are in fact brain states, just like what we thought were witches were in fact normal – or often slightly abnormal – old women. Real witches do not exist. Likewise, there is no such thing as the mental, and if we want to study what we thought were mental processes, we have to study the brain itself. Some philosophers, so-called eliminativists, believe that in the future, we will learn to drop all reference to mental states such as beliefs and desires, and describe ourselves and one another in terms of brain states.

The popularity of identity theory in the philosophy of mind, with its focus on brain states and processes, was linked to the popularity of perceptrons, brain-like computers, in AI in the fifties and sixties. But these perceptrons turned out to be rather unsuccessful. The rise of functionalism in philosophy of mind went hand in hand with the rise of a different computer architecture in AI: that of a program of computations being performed on internal symbols. Yet the idea of brain-like computers got revived in the eighties with so-called neural networks that were quite successful indeed. These were machines with simple units interconnected by connections of different weights, and they were again inspired by the architecture of the brain. Such machines can learn by adjusting the weights of the connections. The deep learning machines of present-day AI are in fact neural networks that can learn their tasks, not by applying logical computations to internal symbols, but by being fed huge amounts of data and changing their connections until they get it right. Whereas in the sixties and seventies identity theory seemed to be replaced by functionalism, on the basis of what was called the first kind of criticism above, the eighties saw a focus on the second kind of criticism on identity theory. Eliminativists or so-called connectionists, encouraged by the success of neural networks, expect that philosophy of mind and psychology will be replaced by neuroscience (Churchland, 1995).

## 5. THE PROBLEM OF CONTENT

Let us now turn to the first of the two main problems in philosophy of mind, the problem of content. The Cartesian problem of interaction was in part the problem of how (mental) thought could act on the (physical) body. And identity theory and functionalism answer that question by saying that thought is itself physical. But how exactly? How can the content or meaning of a thought have causal influence on behaviour? If mental states and processes are, in one

way or another, identical to brain states and processes, isn't it the brain that does all the causal work, and not the content of the thoughts? If my desire for a piece of cheese, causing me to go to the fridge and cut myself some cheese, is in fact a brain state, isn't it the brain state that causes my muscles to lead me to the fridge and not my cheese-desire? For how can a brain state be about cheese? There are roughly three ways of answering this concern.

*The first answer* is to simply say: yes, it is the brain that does all the causal work, and the mental does not do anything because it does not exist. This is the answer of the eliminativists who want to replace philosophy of mind and psychology with neuroscience, and who predict that we will gradually stop referring to thoughts, beliefs and desires at all. This does away with the problem of content, but at the price of being very counter-intuitive. One cannot even say, without contradiction, that they believe that beliefs do not exist.

*The second answer* is to say that mental processes are symbol manipulations, and that these symbols are physical brain states. This answer compares the brain to a computer, which also has internal states that play a role in the causal chain that leads to its output. It is claimed that it is indeed the physical characteristics of these internal states that do the causal work, but that at the same time these internal states do have content or meaning. The idea is that, in a well-programmed computer, the program rules are such that the meanings of the inner states being manipulated are respected, so that the output keeps making sense. If the meaning of one state is "1", and of another "+", the rules are such that "1+1" leads to a state meaning "2". The problem with this answer is that the physical characteristics and the meaning of an internal state of a computer can easily become unstuck. This is because the computer will work whether its internal states have one meaning or another, or none at all. John Searle (1980) illustrates this with his famous Chinese Room argument. Suppose there is a man in a closed room. The room contains a rule book and a slot in the wall. Through the slot, papers are inserted bearing Chinese characters. The rule book, which is in English, contains instructions for reacting to the incoming Chinese characters, which are identified by their form, with other Chinese characters, the output. The man in the room doesn't know any Chinese and doesn't know what the characters stand for. But Chinese people outside the room think they are feeding the room with questions in Chinese and consider the answers they get as output as meaningful and relevant. Searle's conclusion is that, for the computer, it isn't the meaning of its internal states that does any work, it's only their physical form. And just like the man in the room, the computer itself has no idea what it is doing. It is only the interpretation of the users of the machine that gives its states meaning. He thinks this is the difference between computers and human beings: unlike computers, human beings do know what they are doing, and their internal states do have *intrinsic* meaning. It surely seems that Searle is right in

this, but even so he has still to account for *how* internal brain states can have meaning.<sup>6</sup> Thus far, brain scientists and philosophers alike disagree about how to think about meaningful thoughts in all their linguistic, psychological and social complexity, at the level of our brains where all there seems to be are neurons and their connections.

*The third answer* is the one Daniel Dennett defends: there is, actually, no difference between computers and human beings (or human brains). Both are simply physical systems through and through. Both have inner states, but these inner states have no intrinsic meaning. Just as in the case of computers, the inner states of human beings are interpreted from the outside. By whom? Well, by human beings – Searle’s Chinese people outside the Chinese room. But these human beings also have no inner states with intrinsic meaning; these people are also only interpreted from the outside. As Dennett has it: “An implication of the view crudely expressed by the slogan that our brains are organic computers is that just like computers their states can be interpreted [...] by outside observers to have content – and that’s as strong a sort of content as their states can – or could – have. We are both the creators and the creatures of such interpretation” (Dennett, 1982, p. 355). This “interpretationist” answer solves the question how internal brain states can have meaning: they simply do not have any *intrinsic* meaning; all meaning is attributed from the outside and it is all a matter of interpretation. But these claims are problematic: doesn’t one have to have meaningful internal states *already*, in order to be able to interpret someone else’s internal states? And if, as Dennett claims, we interpret *our own* internal states in the same way we do others’, how can we possibly start doing so?

Thus far the problem of content has no answer that is generally accepted.

## 6. THE PROBLEM OF CONSCIOUSNESS

Let us now address the other fundamental problem of philosophy of mind: that of consciousness. Quite suddenly, around 1990, and before there was any definitive answer to the problem of meaning, this other problem became widely discussed in philosophy of mind. How can any purely physical, mechanical thing ever be conscious? Why is it that most of the workings of our brains are completely unconscious, whereas others seem to be conscious or give rise to consciousness? This problem is sometimes called “the hard problem” of philosophy of mind (Chalmers, 1995). Why is there consciousness at all? Couldn’t the brain work perfectly well without there being any consciousness? Couldn’t

---

6. See e.g. Fodor 1987. We know that there are regions in the brain that seem to represent the organism’s own body; actually, there is more than one of this kind of topological neural map. But neural representations of meaningful thoughts are another matter.

there be, for instance, a brain state that fulfills all the alarm-bell functions of pain, without it *feeling* so painful?

Part of the problem is that it is so hard to define what we mean by consciousness. What is it and how can we tell whether some creature or some thing is conscious? Most philosophers agree that it has to do with subjective experience, with what Thomas Nagel famously called a matter of “what it is like” (Nagel, 1974), but there are also philosophers who say that this whole idea of subjectivity, what-it-is-likeness and the qualitative aspect of experience simply doesn’t make sense.

Daniel Dennett (1991) tries to reduce all sensory *experiences* of something to *judgements* that something is the case. If we are looking at a blue curtain, there isn’t *first* an experience of blue and *then* a judgement that the curtain is blue; the judgement “blue there” is *all* that occurs. Thus, he has eliminated the problem of consciousness, and only has to deal with the problem of content, which he solves with the interpretationism explained in the previous section. The problem here is that this solution, though elegant, is *doubly* counterintuitive: we all have the impression that we *do* have sensory experiences, and it is hard to see how one could interpret others without having meaningful inner states of one’s own.

Other philosophers opt for an identity-theory solution: consciousness just *is* a kind of brain state or process, not an effect of it. Pain just *is* the firing of certain neurons. But the search to pinpoint *which* neurons or *what* aspect of neural functioning are identical with consciousness, is still on, with widely diverse answers being given. Moreover, the hard problem of *why* this or that aspect of brain functioning should be conscious remains unanswered. One might think that consciousness must be of some evolutionary advantage, but most scientists agree that consciousness itself doesn’t have any causal efficacy. It is the physical brain states and processes that do the causal work. Sometimes, before I know it, I find myself having eaten another piece of cheese. Apparently, my brain knows perfectly well how to indulge me, without my being conscious of any desire causing my behaviour, or of even the wonderful taste of the cheese. And if so much of brain functioning can be unconscious, why not all of it? Moreover, if consciousness is defined as subjectivity, is it even possible that science, which strives to be as objective as possible, can say anything about it?

As was the case with the problem of content, there is as yet no generally accepted answer to the problem of consciousness.

## 7. MEANWHILE IN SCIENCE 2: THE RISE OF ROBOTICS

Whereas the first decades of AI were dedicated to programming *computers* for doing intelligence-demanding tasks *offline*, without any direct contact between computer and real-life world, in more recent years AI turned towards creating *robots*. Robots behave by definition *online*, interacting with an environment (however artificial it may be). The pioneering work of Rodney Brooks (e.g. 1989) showed that building a robot as a mechanical body with a computer in its head was not the best way of having it navigate through its surroundings. One could build a better one if its legs interacted with the environment relatively independently from one another: robot behaviour was not the execution of a previously computed central plan, but the dynamic interaction of the robot-body parts with its environment.

## 8. THIRD ATTEMPT: EMBODIED EMBEDDED COGNITION AND ENACTIVISM

We have now arrived at the third attempt at dealing with the problems in the philosophy of mind. This third attempt was the result of a quite radical change of direction. Inspired by the progress in robotics, but also by certain insights from other philosophical disciplines, the idea took hold that the focus on the brain that had held sway over philosophy of mind was misguided. The “methodological solipsism” – Jerry Fodor’s term (1980) – of exclusively studying the inner world of the mind-brain, had to be replaced by a consideration of the brain in interaction with the (rest of the) body and the environment. The turn of the century saw the emergence of Embodied Embedded Cognition, which claimed that to study the mental one should consider the brain-body-environment as one dynamic system. For instance, a Japanese child doing sums with the help of an abacus, is not having cognitive processes in her head/brain/mind, which she then turns into instructions for her fingers on the abacus. Instead, it is the whole system of brain and fingers and abacus that does the sums. Cognition was seen to be no longer restricted to processes *in* the brain, but as a process extending beyond skull and skin (Clark & Chalmers, 1998). Likewise, consciousness was no longer seen as a matter of particular brain processes, but as something happening in a whole organism in its environment.

Closely related to EEC, is the view known as enactivism. Already in 1991, the term “enaction” was coined: there is no pre-given world that a pre-given mind internally represents, but mind and world are “enacted” in a dynamic process of interaction (Varela, Thompson & Rosch, 1991). Enactivism was at first an attempt to solve the lingering problem of consciousness. O’Regan and Noë, for instance, claimed that to consciously perceive something was a way of *doing things* (2001). The

difference between experiencing red and green is a difference in the way we act – literally in the way we move our eyes and bodies and know the interdependence between our own movements and the way the world reveals itself. Experiencing what it is like to drive a Porsche is knowing how the car reacts to your own steering and use of the throttle. Perceiving itself was seen as acting, and not as a passive receiving of input from world to mind. Living creatures, as they develop in evolution, cannot fail to acquire this kind of consciousness of their surroundings and their own place within it (Thompson, 2010). This last claim is called the life-mind continuity thesis: the whole universe is just a physical, non-feeling mechanism, but with the origin of life we get consciousness as it were for free. The problem of consciousness is thus solved by claiming that life and consciousness are inextricably linked.

Enactivism and EEC agree that we can no longer approach cognition and the mind more generally as things that are restricted to human brains. It is often thought that these views can best account for *online*, “lower”, cognitive processes, but that some “higher” cognitive processes are “representation-hungry” (Clark & Toribio, 1994). How can one, for instance, do so-called mental arithmetic (no abacus or pen and paper) without keeping some representations of the calculation “in one’s head”? These processes, like abstract problem solving or imagining, can better be explained by postulating an inner domain with content-bearing inner states. Yet there are also philosophers who are not willing to make this admission to the need for representations and go one step further. Radical Enactivists, worried as they are that the problem of content has never been properly solved, claim that no organism has inner content-bearing states, and that no cognitive process is ever in need of inner representations, not even imagining or abstract thinking. *Imagining* the hallway as it used to be, is simply the mentality-constituting interaction of (being disposed to) *walk around* a non-existing table (Hutto & Myin, 2014). For arithmetic they refer to Japanese school children, who start to learn on their abacus, but who are, after training, able to do the calculations with incredible speed *without* the abacus; these children still move their fingers over the non-existing abacus while doing the sums.<sup>7</sup> Of course, tacitly thinking up sentences to say involves meaningful inner states before the words are said out loud. But those meanings are the meanings of *public* language, so the radical enactivists claim, and they are only learnt when a child learns its own mother tongue. Radical Enactivism therefore does not place the great divide of just-physical mechanisms on the one hand, and entities that have something mental on the other, between life and non-life, but between language-users and non-language-users. Although language does play an enormously important part in our lives, it seems odd to claim that prelinguistic children or animals have no

---

7. One may see here echoes of the behaviourist Skinner’s search for covert speech movements during thinking.



mental life at all. But in the Radical Enactivist theory, this does solve the problem of content. With the occurrence of natural language, we get content for free. The problem is thus solved by claiming that natural language is the only thing that has content, and that is for semanticists to worry about.

## 9. CONCLUSION

**A**nd so, philosophy of mind seems to have come full circle: from the simple denial of an inner mental world in behaviourism, to the claim that there are no inner, content-bearing states and that everything is behaviour in Radical Enactivism. Meanwhile, the problems of content and consciousness have led to many different answers, with lots of sophisticated insights, but they still have to be solved. Ignoring the problems, or claiming that they do not exist, hasn't made them go away. Everybody has to pay the piper sometime. Although, amongst philosophers, Descartes' dualism is no option anymore, his idea of an inner 'something' where both thoughts and experiences take place, still, in one way or another, seems to set the agenda for philosophy of mind.

## BIBLIOGRAPHY

- Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. Basic Books.
- Brooks, R. (1989). A robot that walks; emergent behaviors from a carefully evolved network. *Neural Computation*, 1(2), 253–262. <https://doi.org/10.1162/neco.1989.1.2.253>
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 3–34. <https://doi.org/10.1093/acprof:oso/9780195311105.003.0001>
- Churchland, P. (1995). *The engine of reason, the seat of the soul*. MIT Press.
- Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101(3), 401–431. <https://doi.org/10.1007/bf01063896>
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- Dennett, D.C. (1991). *Consciousness explained*. Little, Brown and Company.
- Dennett, D.C. (1969). *Content and consciousness*. Routledge & K. Paul; Humanities P.
- Dennett, D.C. (1982). Comments on Rorty. *Synthese*, 53(2), 349–356. <https://doi.org/10.1007/bf00484909>
- Descartes, R. (2003). *Discourse on method and metaphysical meditations*. (E. S. Haldane & G. R. T. Ross, Trans.). Dover Publications. <https://doi.org/10.5962/bhl.title.44504>
- Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(1), 63–73. <https://doi.org/10.1017/s0140525x00001771>
- Fodor, J. A. (1987). *Psychosemantics*. The MIT Press. <https://doi.org/10.7551/mitpress/5684.001.0001>
- Fodor, J. A. & Lepore, E. (1992). *Holism. A shopper's guide*. Blackwell.
- Hutto, D., & Myin, E. (2014). *Radicalizing enactivism*. The MIT Press. <https://doi.org/10.7551/mitpress/9780262018548.001.0001>
- Levin, J. (2013). Functionalism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2016 Edition).
- Minsky, M., & Papert, S. (1969). *Perceptrons*. MIT Press.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435. <https://doi.org/10.2307/2183914>
- O'Regan, J. K., & Noë, A. (2001). What is it like to see: A sensorimotor theory of perceptual experience. *Synthese*, 129(1), 79–103. <https://doi.org/10.1023/a:1012699224677>
- Park, S. M. (1994). Reinterpreting Ryle: A nonbehavioristic analysis. *Journal of the History of Philosophy*, 32(2), 265–290. <https://doi.org/10.1353/hph.1994.0043>
- Place, U. T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*, 47(1), 44–50. <https://doi.org/10.1111/j.2044-8295.1956.tb00560.x>
- Ryle, G. (1949). *The concept of mind*. Hutchinson.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/s0140525x00005756>

- Thompson, E. (2010). *Mind in life: Biology, phenomenology and the sciences of mind*. Harvard University Press.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: Cognitive science and human experience* (revised edition). MIT Press.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177.

RENÉ DESCARTES



## DESCARTES' SECOND MEDITATION (1641)

---

**D**escartes tries to find a foundation for knowledge. The Scientific Revolution had changed the worldview: the earth was no longer the centre of the universe and the whole world was seen as a mechanical structure. Knowledge could no longer be based on everyday experience. In his search for certainty, Descartes uses the method of doubt: only what cannot be doubted can count as certain knowledge. He cannot be certain that the world around him really exists. Thus, left with just himself, he wants to know what kind of being he is.

In his second meditation, you can see how Descartes starts from old ideas about himself to arrive at a radical new one. First he says what he formerly thought he was: something with a face, hands, arms, that can eat, walk, feel and think. This was the old, common-sense idea of what a human being was. Metabolism, locomotion and sensation belonged to living bodies and only thought belonged to a separate soul.

One thing is remarkable: Descartes speaks of the body as “a mechanical structure such as can be seen in a corpse”. A *mechanical, corpse-like* structure has been substituted for a *living* organism.

Then Descartes starts to doubt the existence of his body and of his eating, walking and feeling. All that is certain is his doubting itself, and that implies that he exists as a doubting, thinking thing.

Feeling remains a problem. If bodies are mechanical, they cannot feel. But surely, he himself is not unfeeling. Descartes argues: feeling may not exist, but *thinking that* one feels does. “What is called 'having a sensory perception' [...] is simply thinking.”

With this step a new dualism of a mechanical, “dead” body and a thinking *and* feeling mind is born.

---

Jenny Slatman

**Plastic Mortals and 21st  
Century Healthcare**

# PLASTIC MORTALS AND 21ST CENTURY HEALTHCARE

## *An Introduction to Philosophical Anthropology*

Jenny Slatman  
[j.slatman@tilburguniversity.edu](mailto:j.slatman@tilburguniversity.edu)

### 1. INTRODUCTION

The first human heart transplant was performed in 1967.<sup>1</sup> Initially, this type of surgery was not particularly successful because recipients usually rejected the donor organ. Nevertheless, this event marked a crucial moment in medicine. With the introduction of cyclosporine, an immunosuppressant drug, in the early 1980s, the transplantation of vital organs became an acceptable and successful medical procedure. Next to the possibility of grafting hearts, it also became increasingly common to open thoraxes to repair hearts beating in a less than ideal fashion. Electric pacemakers now ensure that the heart keeps beating in the right rhythm, and inefficient heart valves are replaced by artificial valves or those from a dead pig. Blocked coronary arteries are bypassed, and here and there a stent can be placed to keep the blood vessels open. People who are waiting for a donor's heart can sometimes even get an artificial heart to bridge the time.

Transplant and implant surgery is not restricted to the heart but can be carried out in many other locations in the body. One might immediately think of kidney and liver transplantation, but there are also stents in veins, arteries, and ducts

---

1. I would like to thank Martine Prange for her constructive feedback on an earlier version of this chapter.

as well as other artificial implants in the joints, teeth, and brains. Additionally, implants are increasingly combined with applications of “smart algorithms,” such as a subcutaneous sensor for people with type I diabetes that can continuously measure the glucose level in the blood, and then automatically adjust it via an insulin pump (see Figure 1). Because hardly anyone will escape having something implanted in their body during their lifetime, we could say that implants, or the possibility of implantation, have become an important part of our existence. Plastic, malleable, artificial people populate our world. Thanks to these technological interventions in medicine, human bodies can be repaired, and human life can be considerably prolonged. Yet, humans remain as mortal as they have ever been.

For philosophical anthropology, the branch of philosophy concerned with the question of how people understand themselves as human beings, the fact that people are aware of their mortality is an important starting point. In classical texts, we can find ideas on how to understand our mortality from the perspective of something immortal, such as the immortal soul in Plato’s *Phaedo*. In our contemporary culture and society, however, the belief in immortality seems to have faded into the background. Our time is mainly characterized by the fact that we try to stretch our mortality as much as possible and postpone our deadlines as long as possible. This stretching has become possible thanks to technological and medical developments.

The emergence and application of all types of medical interventions elicit various philosophical questions. One particular question that often dominates philosophical debate about new technologies in medicine is the ethical one of whether something that is technologically possible is necessarily morally desirable. It remains a question whether “human enhancement”—the application of all kinds of technological interventions to make humans “better than well”—is desirable, or even morally required (Harris, 2010). Very often, discussions on this theme have been hijacked by the so-called transhumanists (those who, like Harris, believe that we should widely distribute technologies that aim at enhancing humans) and bioconservatists (those who believe that we should be skeptical about the use of enhancing technologies because they can threaten human dignity). However, they were in fact philosophically instigated and framed by the Sloterdijk-Habermas debate from around the turn of the century, which explored “therapy” and “enhancement” as a binary pair (Ter Meulen et al., 2007). Whereas therapy is supposed to aim at curing a disease or disorder, enhancement may not be necessary from a strictly medical point of view. Although a neat opposition between therapy and enhancement might be advantageous for various parties, such as health insurers who wish to cover only medically necessary interventions, in reality, the distinction between therapy



and enhancement is not so easy to make due to the ever-changing concepts of health as well as the ongoing tendency of medicalization.

In addition to the broad and general ethical question of whether to embrace the latest technological possibilities, the use of technological interventions has also raised more concrete ethical questions in clinical practice. When treating patients, practitioners always have to consider whether the treatment or intervention is beneficial for the patient, whether it may bring them any harm, and whether their autonomy is respected. From a somewhat broader standpoint, one can also consider whether the treatment is fair. This last point is particularly important in cases of very expensive treatment, for example, or treatments to which not everyone has equal access. Within clinical ethics, these four principles—do no harm (*primum, non-nocere*), autonomy, beneficence, and justice—are central (Beauchamp & Childress, 2001). For each intervention, it is essential to assess whether and to what extent these principles will be acknowledged. If, for example, we look at rather experimental transplantation interventions, such as the first facial transplant in 2005 or the first hand transplant in 1998, one may wonder whether the “do no harm” principle was sufficiently guaranteed. These patients received a treatment for which the long-term consequences were not yet known; the first hand transplant patient ultimately had his donor hand amputated because it was difficult to get used to the new hand (Slatman & Widdershoven, 2010), and the first face transplant patient died ten years later due to cancer caused by the immunosuppressant drugs. Technologies such as the fully automated glucose meter and insulin pump raise justice-related questions. Specifically, these devices are not equally available to everyone because they require a lot of knowledge and (social media) skills: if you would like to manage your own insulin level with the help of an implanted pump, you need to be able to manage a smartphone with the pump’s corresponding app. Philosophy in the form of ethical reflection is therefore always needed to analyze whether the development and use of new medical technologies actually contribute to the “good life.”

Besides ethical questions, the use of new medical technologies also raises other kinds of philosophical questions: Have humans turned into human-machines now that technology has become so prevalent in and on our bodies? Is there still a clear distinction to be made between humans and machines? Further, what about the difference between humans and animals when we consider that animal tissue is used to patch up humans (think of the heart valve of a pig, or the first insulin taken from dogs)? These are the kinds of questions that are the subject of philosophical anthropology. In this chapter, I will explore how these interventions may affect the way we see and interpret ourselves as human beings. I will start with a short description of what philosophical anthropology entails. Since the name “philosophical anthropology” was introduced in German philosophy in the early 20th century and did not travel much further than France

and the Netherlands, it is barely used in the English-speaking world. In my description of the discipline, however, I will make it clear that questions central to philosophical anthropology are inherent in philosophical reflection in general. After this description, I will provide an example of a contemporary philosophical anthropological analysis of intrusive medical technology, for which I will draw on Jean-Luc Nancy's philosophy. In my analysis, I will mainly focus on how the coexistence of human and medical technology has affected our ideas about embodiment and mortality.

## 2. WHAT IS PHILOSOPHICAL ANTHROPOLOGY?

The central question in philosophical anthropology is not an empirical one but rather a reflective one: How do people understand themselves as human beings? Even though such reflection has always been part of philosophy, philosophical anthropology as a separate philosophical discipline has only existed since the beginning of the 20th century. It is no coincidence that anthropology emerged as a separate discipline at that time. The industrial revolution and the rise of capitalism in the 19th century had changed many things at a rapid pace. Many tasks initially performed by humans were increasingly being taken over by machines, and with the rise of medical disciplines such as pathology, physiology, and neurology, medicine was booming. As of the 1920s, more and more philosophers felt the need to reflect on these developments in Western societies. Philosophers who are expressly associated with this tradition include Max Scheler (1874–1928), Arnold Gehlen (1904–1976), and Helmuth Plessner (1892–1985). Philosophical anthropology as a specific philosophical discipline is essentially linked to the German-speaking world.

Other sources of inspiration for philosophical anthropology include the work of 20th century philosophers within the traditions of phenomenology and existentialism, such as Hannah Arendt (1906–1975), Martin Heidegger (1889–1976), Maurice Merleau-Ponty (1908–1961), Simone de Beauvoir (1908–1986), and Jean-Paul Sartre (1905–1980). What all these thinkers have in common is the criticism of the idea that being human is characterized by rationality and reason. The theories of Charles Darwin, Friedrich Nietzsche, and Sigmund Freud—according to Paul Ricoeur (1970), the “three masters of suspicion” (p. 32)—gradually led to the realization that we, as humans, are only small links in evolution, that we are largely determined by existing power relations, and that we are mostly guided by unconscious drives. What the philosophical anthropologists of the first half of the 20th century had in common was that they all underlined the importance of “being situated” and “being embodied” for being human. Below, I will explain in more detail how this focus on embodiment goes hand in hand with a change in thinking about the body (§ 5).

If we assume, however, that the central question of philosophical anthropology is that of what it means to be human, then we immediately see that this strand of inquiry has always been a part of (Western) philosophy in general. If we go back in time and look at the texts of, for example, Aristotle, Plato, and Descartes, different conceptions of what it means to be human emerge. Whereas Plato presents us with the image of the soul that is trapped in a body (*Phaedo*, 81e), Aristotle suggests that a human is a living being endowed with reason, or a *zōon logon ekhon*, or in Latin, an *animal rationale* (*Politics*, 1253a). In the 17th century, René Descartes (1641) interpreted the essence of being human as the combination of the thinking substance (*res cogitans*) and the extended substance (*res extensa*). Julian Offray De La Mettrie (1748), after Descartes, presents an entirely different view of human existence, whereby a human is held to be nothing but a machine. The idea that this machine would also need something like a soul or spirit for cognitive functions is not plausible to De La Mettrie.

These historical conceptions of being human do not just lie behind us. In our times, we often fall back on iconic historical conceptions. For example, in contemporary healthcare, it is apparent that the Cartesian dualistic view of man is still prevalent, and it is for this reason that a strict distinction is made between somatic and mental health problems. In contrast, within contemporary neuroscience and neuropsychology, the materialistic view of humans, which can be traced back to De La Mettrie, is flourishing. Today, many are glad to reduce cognitive and mental functions to brain activity, as this belief is supported by the increasing ability to visualize brain activity (Rose & Abi-Rached, 2013). The conception of humankind that corresponds with this perspective is that of a neuronal machine or network. According to this perspective, humans do not have a non-material dimension. Today, this view is enthusiastically preached by neuroscientists such as Dick Swaab (2014) and Joseph LeDoux (2002).

If we look closely at considerations that support or contradict this neuroscientific-inspired view of humankind, we see that the question of being human—the central question of philosophical anthropology—is also implicitly present within the philosophy of mind. Indeed, while some claim that consciousness can be reduced to the brain or neural activity (Churchland, 1989), others argue that consciousness is not a natural or material phenomenon and thus cannot be explained in terms of science (White, 1991). Since the philosophy of mind is often seen as an analytical or Anglo-Saxon branch of philosophy, we can therefore say that the philosophical movement that deals with the question of being human should not only be seen as a legacy of continental philosophy. Instead, philosophical anthropology, despite its lack of use as a term, is at play wherever philosophers deal with the question of what it means to be human.

Because philosophical anthropology searches for the specific meaning of being human, it also deals with the question of how people differ from Gods, angels,

and animals. In contemporary anthropology, the focus has shifted progressively to the relationship between humanity and technology. When we look at how we interpret our behavior and ourselves, we see that we often mirror ourselves with technological artifacts. Without the developments of steam engines, electrical machines, and computers, we would never say that we “have to blow off steam for a while”; that someone has a “loose wire”; or that “we have to recharge” or even “reset” ourselves. As stated above, the question of how people understand themselves as human beings is a reflective question and not an empirical one, even though this reflection is always nourished by empirical observations. The answer to that question, therefore, is always dependent on the context of the people who are posing it. For this purpose, it is valuable to look at the metaphors we use because they certainly say something about how we describe or interpret ourselves as human beings. By using the machine (or computer) metaphor, we could, for example, explain ourselves as fully material beings controlled mechanically (similar to De La Mettrie’s view), or as dualistic beings if we assume that the machine consists of both hardware and software (Cartesian). These metaphors indicate how we compare ourselves to technology, but technologies also affect our bodies and lives. In the remainder of this chapter, I will focus on how technology may change the way we think about human bodies, our “own” bodies, and our mortality.

### 3. *MORT-VIVANT*

**A**s an example of a contemporary anthropological analysis, I will now zoom in on some ideas of the contemporary French philosopher Jean-Luc Nancy (born 1940). Nancy owes his life to a grafted heart and is therefore in a unique position to reflect on what technology is doing to us as mortals. In several texts, Nancy writes about the meaning of technology in our time, but nowhere do his remarks penetrate as they do in the essay “The Intruder,” a philosophical reflection on the period in his own life in which he suffered from severe heart failure and consequently received a donor heart (Nancy, 2008).

In 1990, Nancy received a donor heart because his own no longer functioned properly. A doctor had told him that his own heart was not programmed to be over fifty years old. For people born in or before 1900, the age of 50 years corresponded to the average life expectancy (CBS, 2019). For someone born in 1940, such as Nancy, 50 is too young to die. Nancy is keenly aware of the fact that he owes his existence to the technological possibilities available in his time. Had he been born with the same heart twenty years earlier, his life may not have been extended because there was no cyclosporine (the immunosuppressant drug that ensures that a donor organ is not rejected) available at that time. As Nancy writes, “I” always finds itself tightly squeezed in a wedge of technical possibilities” (p.

162). His own malfunctioning heart, and the ability of technology to fix it, invited him to think about what it means when we talk about our “self” or “I,” or when we speak of “my” body.

Nancy indicates that there is an ongoing process between what we call our “own” or “self” and what we call “strange” or “other.” In his essay “The Intruder”, he describes the different ways in which “strangeness” can present itself. First, there is the sick, malfunctioning heart. Where at first you are never aware of your own heart—which in itself also indicates a strangeness toward yourself—you later become aware of your own heart in a negative way, for example, through chest pain, shortness of breath, or nausea in your mouth: “I had this heart at the tip of my tongue, like improper food” (p. 163). Nancy’s heart had alienated itself from his own body. Thus, to save his own body and self, the sick heart had to be taken away and replaced by another.

Then there is the donor’s heart, or the heart of another person, which may be of a very different age than you are. Moreover, it could be the heart of someone of a different gender or ethnic origin. To be able to accept that strangeness in your body, your own immune system has to be suppressed. How paradoxical this idea is: to protect one’s “self,” the defense mechanism must be put on inactivity for that same “self” to survive. For Nancy, lowering his own resistance meant that other uninvited intruders, such as viruses, could go about their business more easily. Also, treatment with cyclosporine left him with lymphatic cancer. Cancer, he writes, “is like the ragged, crooked, and devastating figure of the intruder” (p. 168). Cancer can be construed as an intruder, but, at the same time, it is our own cells that turn against ourselves. Chemotherapy, radiotherapy, and stem cell transplantation—all therapies that target both the malignant and the healthy tissue—ultimately had to provide a livable situation for Nancy.

Nancy describes how his own “self” continues to exist thanks to strangeness, strangers, or *Fremdkörper*. In its ultimate form, the strange nature of *Fremdkörper* involves the peculiar nature of death. To us, as living individuals, death represents strangeness *par excellence*. After his heart transplant, Nancy’s son characterized him as “mort-vivant”: a living dead person (p. 170). In multiple respects, this adequately describes the situation. With a heart that is so diseased that its replacement is crucial for survival, one is first bound to die. It is abnormal to live on when having a heart “programmed to live fifty years.” One lives while one should have been dead already. However, “mort-vivant” also refers to the dead donor body. You live because someone else has died. The one heart, donated and accepted, connects the dead with the living.

Although Nancy does not call himself a philosophical anthropologist, I would say that his analysis of his own heart transplantation is typical philosophical anthropology. His analysis revolves around the question of what it means to say

“I” or “one’s own,” and to that end, he is discussing the contemporary phenomenon of heart transplantation. His analysis and questioning of being human are thus very much based on the contemporary context, and even more specifically, on the context of his identity as the philosopher Nancy, who is a heart patient in contemporary France. As such, at first sight, his analysis resembles a medical anthropological study. In medical anthropology, empirical research methods, such as ethnography and interviews, are used to investigate the impact that certain medical practices have on a particular population. However, Nancy’s analysis goes further, as it entails more than an auto-ethnography that indicates the medical practice of cardiology in France in the 1990s. With his historical and contextual analysis, Nancy wants to reflect further on the meaning of being human in a more general sense. For example, he says that the heart transplantation acted as a kind of eye-opener. What he means is that the possibility of heart transplantation enables us to say something about the relationship between humanity and technology in general (in our time) and about the relationship between what is “own” and what is “strange” in ourselves. These two factors are inextricably linked. There is no human being without technology, and there is no such thing as a self apart from something strange or other.

#### 4. MORTALS FOREVER

**E**ven though the industrial revolution of the 19th century launched humankind into the technological era, technology has always been part and parcel of human life and contributed to making life easier and postponing death. Our forebears went out to look for food carrying clubs and stones. This is what Helmuth Plessner refers to when he claims that people are artificial by nature, as we have always used artifacts. Nancy repeats this idea when he writes that humans have always been the most terrifying and troubling forger who “denatures and remakes nature” (p. 170). In our era, however, technology has received yet another meaning. Contemporary technologies invite us to reflect on our own mortality in another way, and it is their invasiveness and omnipresence in our lives that shows us that the concept of a natural human being, one who is opposed to technology, has become a naive idea.

The most characteristic strangeness of our time is not the bizarre nature of technological feats, but our sustained effort to defer death. If we look at the averages of life expectancy, we can see that we have succeeded in mitigating death wonderfully well. If you were born in 2000 in the Netherlands or a similar European country, your average life expectancy was over 80 years. Because this is an average, and because people with a higher level of education—people who read this book—are surviving beyond their expectancy, there is a good chance that you will live more than 90 years, whereas the average life expectancy for

someone born in 1900 was only 50 years (CBS, 2019). This means that through technology (which includes the construction of good sewers and clean water drains, the development of effective vaccines and drugs, rapid innovations within surgery, and the automation of many processes), we have gained nearly 40 years of life per individual in a century. This dramatic demographical shift has certainly had repercussions on the way our society is organized. However, Nancy raises another crucial point by claiming that contemporary technology has rendered death more visible than ever, for “to defer death is also to exhibit it, to underscore it” (p. 165). In his view, technology has taken over the role of religion in Western societies. If religion offered us a vision of the infinity of life after death, technology merely promises a postponed death. Nancy thus claims that technology emphasizes our finite nature, because by wanting to postpone death for longer and longer, we no longer seem to expect that there is still something ahead of us beyond our own finitude.

## 5. THE BODY'S PLASTICITY

What is striking about Nancy's text is that when he talks about the “I” or “self,” he is not talking about some kind of thinking substance or mind. The “I” or the “self” is nothing but our embodied existence. Nancy's thought here is closely related to that of the philosophy of existence and phenomenology of the first half of the 20th century. As indicated above, the strongest characteristic of this group of thinkers was that they no longer perceived human beings as an *animal rationale* or a dual being in which the body is guided by the mind. For them, being human is primarily determined by embodiment. It is important to note that this criticism of the idea of the *animal rationale*, or more specifically of Cartesian dualism, goes hand in hand with another philosophical view of the body. In Descartes's view, the body is a *res extensa*, or an extended thing. For Descartes, the human body is thus similar to any other thing or object that takes up space. This idea of the body as a thing, or also a machine, has been confirmed and strengthened by the developments in medicine from the 18th century onwards, from which anatomy was introduced into clinical practice. Whereas in the past, illness was seen as a general imbalance of bodily fluids, from the 18th century, people started to look for the cause of a disease within a certain place in the body. The bodies of people who had died of a particular disease were dissected to look for specific defects in their anatomical structures. This clinical and epistemological development marks, according to the French philosopher Michel Foucault (1973), the birth of the “medical gaze.” In this way, the body changes from a whole that needs to harmonize with itself and its environment into an anatomical-physiological object that can be dissected and examined with millimeter precision.

If we go back to the texts of philosophical anthropologists at the beginning of the 20th century, it becomes evident that they find this specific view of the body—the body as a thing or object—far too limited. In the German texts by, for instance, Scheler, Plessner, and Husserl, we find an addition to this idea in their indication of a distinction between *Körper* and *Leib*, which is difficult to translate into English. *Körper* refers to the body as a thing: your own body when you attribute specific physical qualities to it, and the body that the doctor examines, but also the body that the undertaker puts in the coffin. It is indeed very common to take the bodies of others, as well as your own body, as a thing, an object, or *Körper*, and to experience it as such. In doing so, you take an external perspective on the body as you look at it from the outside. However, according to Husserl, and this idea was later adopted by Merleau-Ponty, you can also experience your own body from a first-person perspective. In this case, you experience your body from within. This experience is formed by sensations that are localized in the respective sense organ (Husserl, 1952, § 36), including the sensations of touch, pain, warmth or cold, and even your posture and your own movement. Husserl calls this sensation of your own body the *Leib* experience. In this experience, your body does not appear as an object, as it does in the *Körper* experience, but instead, it is perceived as an embodied zero point, or an embodied here and now from which you act, perceive, and can orient yourself. Merleau-Ponty calls this zero point the “embodied subject” or the “lived body” (*corps vécu*; Merleau-Ponty, 1945).

With the idea that the body is not only an object but also a subject, there is a definitive departure from the classical idea that the subject corresponds to something immaterial like the mind. Subjectivity is not something that can be cut loose from embodiment. This is why Nancy’s “I” must be understood as an embodied self. In his analysis of the meaning of embodiment, Merleau-Ponty shows that the boundaries of the lived body do not simply coincide with the body as a thing or object. The lived body, which Merleau-Ponty also calls “one’s own body” (*corps propre*) given its foundation in the first-person perspective, is a dynamic whole. The boundaries of our own bodies are partly formed by the degree to which things can be incorporated. When something is incorporated, that thing also becomes part of our lived and embodied zero point. Here you might think of the use of tools or instruments. Merleau-Ponty himself gives the example of a blind man with a white cane. When the man uses the cane to explore the world and is no longer aware of the cane as such, the tool has become an extension of his body. It is then incorporated, or completely appropriated (Merleau-Ponty, 1945, p. 144).

This incorporation also takes place when people who use a prosthesis are no longer aware of the artificial device as such. When used flexibly, the prosthesis is no longer experienced as an object but instead becomes a part of the embodied



acting zero point. Incorporation is largely based on habituation and can be understood as the appropriation of something that was previously strange, as the appropriation of *Fremdkörper*. As indicated above, we are very well able to continue living with all kinds of artificial implants and with other people's organs and tissues. Hence, the boundaries of what we call our own body seem fairly elastic. The idea that the boundaries of one's own body are not fixed in advance has become an important theme in contemporary feminist philosophy. For example, Haraway (1992) uses the term "cyborg"—a combination of organism and cybernetics or self-regulating technologies—to indicate that it is not so easy to distinguish humans from technology. In addition, Haraway claims, the distinction between humans and animals is not so clear when we think of cases like the so-called OncoMouse (a laboratory-produced rodent with an oncogene) that is used to test cancer treatment. The human as a cyborg is thus human, animal, and technology (Haraway, 1992). Within feminist philosophy, this blurring of boundaries, which goes hand in hand with the elastic boundaries of one's own body, is by no means considered a loss of humanity. Instead, it invites us to reconsider the idea that identity formation is solely determined by nature or biology, and as such, it contributes to overcoming entrenched gender patterns. The blurring of boundaries emphasizes that what and who we are is not so stable, and that (gender) identity is flexible and multiple rather than fixed and unambiguous (Braidotti, 2013; Grosz, 1994).

In considering Nancy's work against the backdrop of feminist philosophy, it is evident that he is less concerned with issues pertaining to (gender) identity politics, even though some claim that his ideas on embodiment are of great value for feminism (Perpich, 2005). As I see it, his work provides a more general criticism—broader than the feminist criticism—of the idea of our "own body" and the very idea of the "self." A clear difference between Merleau-Ponty and Nancy is evident here. Whereas Merleau-Ponty emphasizes the experience of "ownness" that arises when something is incorporated, Nancy emphasizes that this experience of "ownness" is always related to a form of strangeness. Based on Merleau-Ponty's phenomenology, we could say that in a successful operation, the donor heart or an artificial implant has become part of one's own body. On the contrary, Nancy states that there can be no own body without instances of strangeness. This can be understood as follows: the experience of the body as one's own, the *Leib* experience, is only possible if at the same time we assume that the *Leib* also has a certain thingness, or that the *Leib* is also *Körper* (Slatman, 2014). Without the-being-extended of a thing, the *Leib* could never be touched, either by itself or by something outside of itself, and it would then have no experience of itself. If we interpret the *Körper* as that which in a certain sense is strange to ourselves because we can distance ourselves from it, then it becomes clear that our own body always already has something strange

within it. We as humans are capable of incorporating strange elements into our own bodies solely because our own bodies already have something strange.

What a phenomenological anthropological analysis of contemporary technologies teaches us is that technology does not only involve an addition to or transformation of our human biology. The many possibilities of today's technologies hold up a specific mirror that allows us to interpret ourselves in a certain way. As long as the dream or science fiction of human immortality cannot be realized by technology, then technology will only postpone death. We succeed so effectively in delaying death because our bodies are so plastic and thus capable of absorbing strange elements. Looking at the possibilities of technology, we, contemporary humans, can interpret ourselves as *plastic mortals*. More than ever, as Nancy underlines, we are aware of our finiteness because the growing trust in technology seems to go hand in hand with a waning belief in an afterlife. More than ever, we are aware that our own being can never be entirely our own, and that we are always embedded in a network of strange elements. Being human means that one relates to what is strange in a pliable way, and in many cases, this is literally associated with strange material in one's body, such as plastic.

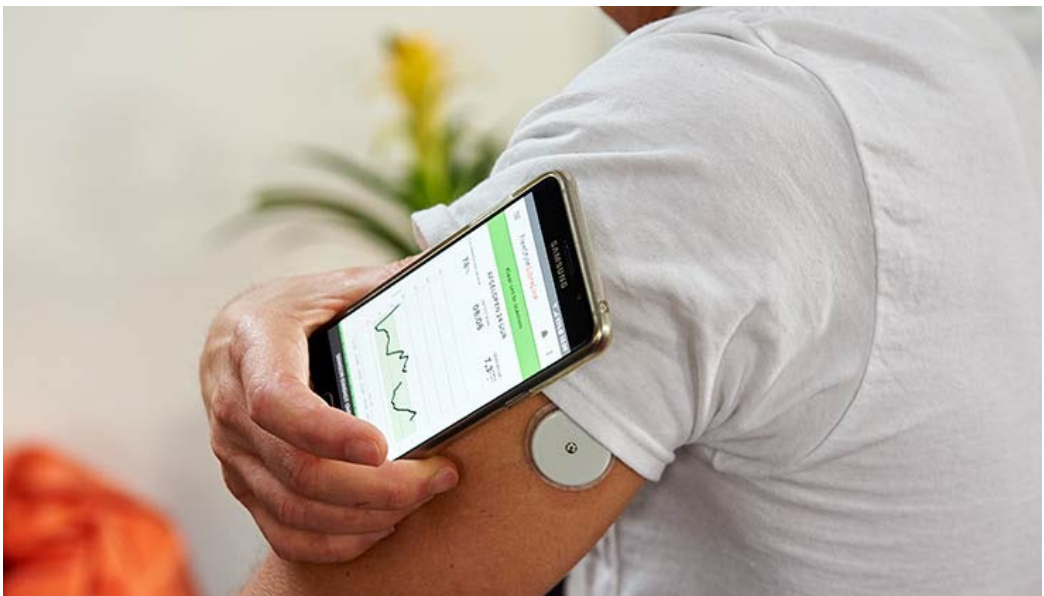
## 6. PHILOSOPHICAL ANTHROPOLOGY FOR 21ST CENTURY HEALTHCARE

**W**hen we ask how people understand and interpret themselves as human, then it is virtually impossible not to discuss the relationship between humans and technology. Therefore, I believe that in our time of increasing technological innovations, the philosophical discipline of philosophical anthropology is of great importance. Technological developments to improve the human body, such as those I have described in this chapter, will only multiply. In addition, more and more technologies will be used to boost human cognition. Already for a long time, computers have been much better than humans in making calculations and playing chess. Recently, due to the great success of deep learning in artificial intelligence (AI), the machine has also beaten humans when it comes to interpreting pictures. AI is now also conquering the medical world, and there are already numerous applications within the healthcare sector (Topol, 2019). The subcutaneous sensor together with the insulin pump (see Figure 1), which helps people with type I diabetes manage their blood glucose levels (which I spoke of in the introduction), is also guided by a deep-learning program. This technology-based system becomes increasingly familiar with the user's body, and it reacts to fluctuations in ways more precise than the user has been able to do herself or himself. The use of such a "smart" technological device brings a new dimension to the coexistence of humans and technology. For indeed, this technology not only involves the implementation of a *Fremdkörper*. The artificial

intelligence embedded in this device will also boost the user's knowledge and awareness of their body. This new type of "self-knowledge" can also be subject for philosophical anthropological scrutiny.

In the face of all kinds of technological developments, philosophical anthropology can ask what exactly is going on, and in what ways the portrayal that humans have of themselves can be changed by that technology. In that capacity, philosophical anthropology is descriptive and not prescriptive: it will not directly address whether the new technology should be used. Herein lies also the difference with an ethical approach, as I explained in the introduction. Nonetheless, a philosophical anthropological approach can also have practical purposes. When we have a better understanding of our own embodied being and how we incorporate technologies, we can also apply that knowledge in (health) practice. A good example of this is the development of a tool to determine the Patient Transparency Diagnosis, which would measure the extent to which an implant is felt by a user (Tbalvandany et al., 2019). Such a tool can help to make the coexistence of people and technology as pleasant as possible. When philosophical anthropologists seek to provide a practical benefit to healthcare, it is recommended that they not only conduct their philosophy from their armchair, but that they also become thoroughly acquainted with the field they are analyzing.

Figure 1



Subcutaneous sensor and insulin pump. Picture by courtesy of Abbott company.

## BIBLIOGRAPHY

- Aristotle. (1995). *The Politics of Aristotle* (E. Barker, Trans.). Oxford University Press. <https://doi.org/10.1093/actrade/9780199538737.book.1>
- Beauchamp, T., & Childress, J. (2001). *Principles of biomedical ethics*. Oxford University Press.
- Braidotti, R. (2013). *The posthuman*. Polity Press.
- CBS. (2019). Levensverwachting; geslacht, leeftijd (per jaar en periode van vijf jaren) [Life expectancy; gender, age (per year and 5 years)]. *CBS-Statline*. <https://doi.org/10.1093/gmo/9781561592630.article.53234>
- Churchland, P. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. The MIT Press. <https://doi.org/10.7551/mitpress/4952.001.0001>
- De La Mettrie, J. (1996). Machine man. In A. Thompson (Trans.), *Machine man and other writings* (pp. 1–40). Cambridge University Press. (Original work published 1748)
- Descartes, R. (2008). Meditations on first philosophy. In M. Moriarty (Trans.), *Descartes: Meditations on first philosophy, With selections from the objections and replies*. Cambridge University Press. (Original work published 1641)
- Foucault, M. (1973). *Birth of the clinic* (A. M. Sheridan, Trans.). Routledge. (Original work published 1963)
- Grosz, E. (1994). *Volatile bodies: Toward a corporeal feminism*. Indiana University Press.
- Haraway, D. (1992). *Simians, cyborgs and women: The reinvention of women*. Free Association Books.
- Harris, J. (2010). *Enhancing evolution: The ethical case for making better people*. Princeton University Press. <https://doi.org/10.1515/9781400836383>
- Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy, second book*. Kluwer Academic Publishers. [https://doi.org/10.1007/978-94-009-2233-4\\_12](https://doi.org/10.1007/978-94-009-2233-4_12) (Original work published 1912)
- Ledoux, J. (2002). *Synaptic self: How our brains become who we are*. Viking Penguin
- Merleau-Ponty, M. (2012). *Phenomenology of Perception* (D. A. Landes, Trans.). Routledge. (Original work published 1945)
- Nancy, J.-L. (2008). The intruder. In R. A. Rand (Ed.), *Corpus* (pp. 161–170). Fordham University Press. <https://doi.org/10.2307/j.ctt13x04c6.9>
- Perpich, D. (2005). Corpus meum: Disintegrating bodies and the ideal of integrity. *Hypatia—a Journal of Feminist Philosophy*, 20(3), 75–91. <https://doi.org/10.1111/j.1527-2001.2005.tb00487.x>
- Plato. (1997). *Complete works* (J. Barnes et al., Ed.; J. M. Cooper, Trans.). Hackett.
- Ricoeur, P. (1970). *Freud and philosophy: An essay on interpretation*. Yale University Press.
- Rose, N., & Abi-Rached, J. M. (2013). *Neuro: The new brain sciences and the management of the mind*. Princeton University Press.
- Slatman, Jenny. (2014). *Our strange body: Philosophical Reflections on identity and medical interventions*. Amsterdam University Press. <https://doi.org/10.1017/9789048523146>

- Slatman, Jenny, & Widdershoven, G. (2010). Hand transplants and bodily integrity. *Body & Society*, 16(3), 69–92. <https://doi.org/10.1177/1357034x10373406>
- Swaab, D. (2014). *We are our brains: A neurobiography of the brain, from the womb to Alzheimer's*. Spiegel & Grau
- Tbalvandany, S., Harhangi, B., Prins, A., & Schermer, M. (2019). Embodiment in neuro-engineering endeavors: Phenomenological considerations and practical implications. *Neuroethics*, 12(3), 231–242. <https://doi.org/10.1007/s12152-018-9383-6>
- Ter Meulen, R., Nielsen, L., & Landeweerd, L. (2007). Ethical issues of enhancement technologies. In H. D. R. E. Ashcroft, A. Dawson & J. R. McMillan (Ed.), *Principles of health care ethics* (pp. 803–809). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470510544.ch108>
- Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.
- White, S. (1991). *The unity of the self*. MIT Press

### RECOMMENDED FURTHER READING

- Slatman, J. (2016). Is it possible to “incorporate” a scar? Revisiting a basic concept in phenomenology. *Human Studies*, 39(3), 347–363. <https://doi.org/2010.1007/s10746-10015-19372-10742>
- Slatman, Jenny. (2019). The Körper-Leib distinction. In A. Murphy, G. Weiss, & G. Salamon (Eds.), *50 concepts for a critical phenomenology* (pp. 203–210). Northwestern University Press. <https://doi.org/10.2307/j.ctvmx3j22.33>

SIMONE DE BEAUVOIR



## MAURICE MERLEAU-PONTY (1908-1961)

### *PHENOMENOLOGY OF PERCEPTION*

---

**M**erleau-Ponty is usually seen as the philosopher of the body and embodiment. German philosophers, such as Plessner and Husserl, had already indicated that the idea of the body as a thing or object (*res extensa*) is too limited and that the body is not only *Körper* (objective body) but also *Leib* (lived body). In Merleau-Ponty's main work, *Phenomenology of Perception* (1945), the lived body is explained in terms of "one own body" (*corps propre*) or "embodied subject" (*corps sujet*). For the first time in the history of modern Western philosophy, the body is conceived as a subject. As a French philosopher educated in the 1930s, Merleau-Ponty is particularly influenced by German phenomenology (Husserl) and philosophy of existence (Heidegger). Where Husserl speaks of sense-making (*Sinngebung*) in terms of intentionality or "consciousness of," Merleau-Ponty makes it clear that before we are explicitly aware of anything we already give meaning through our embodied actions. He, therefore, speaks of "motor intentionality." Whereas Heidegger describes human existence as *Dasein*, Merleau-Ponty underlines that existence always involves being embodied. It is striking that Merleau-Ponty also had great knowledge of the psychology, neurology, and psychiatry of that time. Especially the work of the German psychiatrist and neurologist Kurt Goldstein who, among other things, examined and treated World War I veterans, was very important to him. Merleau-Ponty's analyses of what it means to be human are thus developed in close dialogue with empirical findings from that time and are not only based on philosophical theories. Jean-Luc Nancy (born 1940) nowhere explicitly discusses Merleau-Ponty's work and does not call himself a phenomenologist because he finds this movement too subject-centered. Nevertheless, his thinking does fit within the philosophy of existence (following in the footsteps of Heidegger). And just like Merleau-Ponty, he describes existence as embodied. Unlike Merleau-Ponty (and also unlike Heidegger), Nancy emphasizes that the (embodied) subject is not the center of sense-making, but that sense always emerges in the "between," the "being with" of multiple bodies.

---

Bart Vandenabeele

**Appreciating  
Beauty and Art**



# APPRECIATING BEAUTY AND ART

## *An Introduction to Aesthetics*

Bart Vandenabeele  
[bart.vandenabeele@ugent.be](mailto:bart.vandenabeele@ugent.be)

### 1. INTRODUCTION

**N**ot only artists and art connoisseurs, but also web designers, gardeners, carpenters, photographers, musical directors and managers make aesthetic judgements. When decorating our house, developing a logo for a firm or inviting friends over for an evening meal, we make all kinds of decisions that are based on aesthetic issues, that is, issues concerning the way things look or fit in. Which tablecloth will I use tonight, shall I put candles on the table or not, which clothes shall I wear, and so on. In our everyday choices, aesthetic values thus play an important part, even though we may not always realise this ourselves.

Moreover, when we are shopping for clothes, visiting a museum or attending a play, we often discuss our aesthetic judgements with friends and relatives. This can be a highly sensitive matter, as our aesthetic taste may reveal something about our personality in displaying what we value in life. People may therefore sometimes be reluctant to discuss their aesthetic preferences, being sceptical about whether aesthetic judgements can be compared, assessed and disputed. This raises the question whether it is at all possible to ground our aesthetic appreciations rationally. Can philosophers who are experts in aesthetics and philosophy of art offer relevant insights here?

What is the philosophical discipline of aesthetics about? One of the highly debated issues is whether aesthetic appreciations are subjective or objective. On the one hand, an aesthetic judgement such as 'This sunset is beautiful' seems to be merely reporting a subjective feeling of the person who utters it. On the other hand, as the 18th century philosopher Immanuel Kant argues, we are 'suitors for agreement' and enjoy sharing our aesthetic preferences. At least some types of aesthetic appreciations seem to lay claim to interpersonal or intersubjective validity, i.e., to require that others concur, and judgements of the type that some x is beautiful seem to be valid candidates.

Yet the traditional focus of philosophical aesthetics on beauty has become undermined by the practice that used to be a primary object of aesthetic study, namely art. Since the rise of abstract, non-figurative, conceptual art and the abundance of Duchamp-style readymades (prefabricated ordinary objects, such as a urinal or a snow shovel, that are taken into a museum and are thus elevated to the status of art by the artist), the art world has turned itself away from beauty and sometimes even aesthetic value altogether. This constitutes a challenge to philosophical aesthetics and philosophy of art, for the aesthetic value can no longer claim to be the only or even primordial value of the artefacts that we consider to be genuine works of art. Before considering the complicated issue of the values of art, we now first turn to questions about the meaning and status of aesthetic appreciation, more specifically the judgement of beauty and its purported intersubjective validity.

## 2. THE JUDGEMENT OF BEAUTY

As indicated, the philosophical discipline of aesthetics is not confined to exploring art or artistic value, but it is also a study of aesthetic judgement or appreciation, i.e., of the way in which we contemplate, experience and assess objects from an aesthetic point of view. In other words, aesthetics investigates what has traditionally – since the birth of the discipline in the eighteenth century – been called *the judgement of taste*. What kind of judgement is this? A judgement of taste is not merely a statement of personal preference such as 'I like vanilla ice cream' or 'I hate cauliflower', but an attempt to express what is aesthetically worthwhile in a particular (natural or artistic) object. Yet there is no agreement on what it exactly means to say that something is 'aesthetically worthwhile' or how to assess an object's aesthetic qualities.

The term 'aesthetic' is derived from the Greek word 'aisthèsis', which means sensory perception or sensation. Literally speaking, aesthetics would thus be the 'doctrine of sensory perception'. Alexander Baumgarten (1714-1762), who is usually credited to be the first to use 'aesthetics' in its modern sense, calls the discipline of aesthetics the study and perfection of sensible cognition (*cognitio*

*sensitiva*). As Baumgarten insists, aesthetic cognition is not inferior to intellectual cognition: it is not an inferior and provisional substitute for logical and scientific reasoning, but a worthy equivalent of it. What we perceive through aesthetic cognition is not merely the (lack of) perfection of an object. It is also the use of a broad range of our mental capacities that guides our exploration of sensible representations or imagery and our contemplation of the pleasing aspects of an object that offer us an experience of aesthetic pleasure. Baumgarten was the first to define the aim of aesthetics as the perfection of sensible cognition (*perfectio cognitionis sensitivae*), which is beauty. On Baumgarten's view, what may ground aesthetic pleasure is not merely perceiving the perfection of an object (so not merely the representation of some objective perfection in a form accessible to the senses), but also exploiting, developing and refining the possibilities of sensible representation for their own sake (see Guyer 2014, 326-329). Thus, aesthetic pleasure, and hence beauty, may be produced by perceiving perfection through the senses as well as by perfecting our sensible and imaginative capacities. Of course, sensory perception may be pleasurable without offering beauty. The smell of the petunias in my garden, my wife's perfume, the sensation of sinking into a hot bath, that gorgeous meal in my favourite restaurant, and so on: these are all pleasurable sensations, but experiencing pleasurable sensations is not sufficient for appreciating and experiencing beauty.

In addition, not all pleasure we take in beautiful objects is (primarily) sensual: surely the pleasure of enjoying a novel or poem is not sensual. Moreover, not all our senses seem fit to appreciate beauty. The senses of taste, smell and touch do not seem to be our usual tools for sensing beauty. It is plausible to think that the relevant senses for appreciating beauty are sight and hearing, though it is not easy to categorise every experience of beauty as one of these two. Consider, for instance, the reading of a novel: to appreciate the beauty of, say, Tolstoy's *War and Peace* we need our eyes, but it is not the letters as such of course that are beautiful; the beauty of the novel is based on the meanings of the words and their relation to the way the author expresses them.

What is the difference, then, between sensual pleasure and the pleasure of beauty? Beauty requires more than merely agreeable sensations. Immanuel Kant, who provided the first elaborate account of the logical requirements of the judgement of taste, distinguishes between the agreeable and the beautiful. Judgements of the agreeable, expressed by saying that one likes (or dislikes) something or finds it (dis)satisfying (for example, food, drink, perfume ...) are statements of merely personal preference. In contrast, judgements of the beautiful are *normative* judgements that make a claim to universal validity, i.e., they are presented as binding on everyone. For, as Immanuel Kant argues, the judgement of beauty is based on disinterested pleasure (*Wohlgefallen*) and not on merely sensual excitement or satisfaction (*Vergnügen*).

Furthermore, beauty concerns pleasure *at* something. Rather than being primarily about sensations within ourselves (as when enjoying a cool morning breeze on our skin), the pleasure is directed outwards. It is about the beautiful *object*, instead of about me: it implies, as it were, a joyful affirmation of the presence of, say, a beautiful landscape, painting or sculpture. For this reason, one may call this kind of aesthetic pleasure *contemplative*: beauty is a reason for attending carefully to the particular object that possesses it. The term ‘aesthetic’ is therefore somewhat misleading, as it suggests that what we consider as aesthetic pleasure is always necessarily connected to sensory pleasures (as Baumgarten suggested). Yet – as George Santayana aptly puts it in his influential essay on *The Sense of Beauty* (1896): “beauty is pleasure regarded as the quality of a thing” (Santayana 1955, §11, 31).

Judgements of beauty are taken to be the privilege of rational creatures, who possess imagination, for only human beings are capable of enjoying particular objects for their own sake and providing reasons for their aesthetic appreciations. For instance, when asked “*Why is this sunset beautiful?*”, human beings are supposed to be able to give reasons that are related to the qualities of the very sunset which is the focus of the aesthetic judgement. Hence, although appreciating beauty is not based on a description of the object, but on a personal feeling of pleasure, beauty is not *merely* subjective. Contrary to sensory pleasures (‘sensations’) or judgements about the (dis)agreeable features of something (e.g. spinach or oysters), a judgement of beauty such as “This sunset is beautiful” claims, as it were, to disclose an ‘objective’ quality of the aesthetic object, for instance this sunset that I am admiring now in Venice. The question therefore arises whether the judgement of beauty is at all possible, i.e., whether a judgement exists that can be subjective (and hence relative) as well as objective (and hence intersubjectively valid). This has been one of the main questions of philosophical aesthetics since the eighteenth century: How can an intersubjectively valid form of judgement be based on something so personal as the pleasure felt in appreciating a particular object as beautiful? It is to this complex issue that we now turn.

### 3. BEAUTY: THE PROBLEM

**A**ppreciating beauty is founded upon a personal feeling of, what Kant called, disinterested pleasure. Yet, the pleasurable feeling is not *merely* personal, hence the word ‘disinterested’: I also demand that others ought to feel the same, that is, that they, in their disinterested feeling, ought to concur with my judgement. (Whether they will agree or not is an empirical matter.) When I admire a beautiful landscape painting by Constable, for example, I do not merely experience agreeable sensations in myself. My aesthetic appreciation (“This

painting is beautiful') does not report on inner sensations, but expresses that the painting *is* beautiful. A judgement of taste does not describe my personal feeling of pleasure, even though it is based on it.

This however looks contradictory; it leads to what Kant technically refers to as an *antinomy*, i.e., a contradiction which he believed inescapably follows from our rational nature. The antinomy of the judgement of taste can be formulated as follows: this type of judgement is grounded in a subjective feeling of pleasure (or displeasure in the case of negative judgements of taste) *and* lays claim to objective validity, since it is presented as binding on everyone. Whereas the beauty of a thing cannot be proved by means of objective rules or concepts, I can reasonably be asked to justify my appraisal and we can argue about the beauty of the thing. There can actually be a genuine *disagreement* (and not simply a *difference* as is usually the case in, for instance, disputes about the tastes in food) about whether, for instance, Franz Schubert's song-cycle *Die Winterreise* is really more beautiful than Gustav Mahler's *Lieder eines fahrenden Gesellen* or not. I would definitely say that the former is more beautiful than the latter, but you may not agree.

One cannot provide objective principles that determine once and for all what is beautiful and what is not, or whether something is more beautiful than something else. However, beauty is not mere sensual attraction. We can (attempt to) explain why Jane Austen's novel *Pride and Prejudice* is more beautiful than *Mansfield Park*, because (for instance) the description of the characters is subtler and more penetrating in the former, or why the film *The Godfather II* is better than *The Godfather III*, because the former's plot is more balanced and the main character possesses more psychological profundity. Thus, even though there may not be objective principles that determine beauty, there may well be rational grounds for appreciating it.

#### 4. IS TASTE RELATIVE?

Since no objective principles seem to be available to determine the correctness of aesthetic appreciations, a number of thinkers, especially so-called postmodernist and neo-Marxist ones, contend that art has no specific value or that the value of art differs hardly or not at all from non-artistic artefacts, such as cheap soap series, vulgar B-films or cheesy potboilers. They argue that 'taste is relative' and that the distinction between art and non-art, between high and low culture or beauty and kitsch is ultimately an 'ideological' construction, which serves merely to maintain certain elitist ideals and 'bourgeois' institutions. Furthermore, they contend, these ideological concepts and distinctions blind us to the social truth. Thinkers such as Michel Foucault, Gilles Deleuze, Pierre Bourdieu, Terry Eagleton and several others argue (customarily in a very flamboyant way)

that universal values do not exist and that the Western canon of great art and literature is actually delusional, authoritarian and unjust. Only a struggle against the 'bourgeois' values can overturn the canon and liberate us from the chains of 'power'. If these thinkers are right, this has radical consequences for the very concept of beauty, since the concept could no longer be justified, except as an 'ideology', that is, as a theory that is accepted merely for its (pernicious) political and social utility and not for its truth.

This postmodernist critique of the 'ideological' nature of beauty contains a number of valuable elements.

First, this critique draws attention to the social and political injustices that are downplayed by those philosophers and art critics that focus exclusively or primarily on the aesthetic merits of artworks. For instance, downplaying or ignoring the social, moral and political outrage at the horrors of war that is conveyed by superb masterpieces like Goya's *The Third of May, 1808* or Picasso's *Guernica*, while only focusing on the fine rendering of the depicted scenes can hardly be called a sensible and fair assessment of the value of such artworks. To focus only on the aesthetic qualities of such works is one-sided and displays a lack of understanding of what the paintings are really about and what turbulent emotions are expressed in them.

Second, the above-mentioned thinkers and their followers criticise the 'Eurocentric' one-sidedness of focusing on so-called Western art and aesthetics, and demand the recognition of non-Western artistic and aesthetic traditions.

Third, this type of critique emphasises the provisional and context-bounded character of aesthetic and artistic appreciations. Those are inevitably influenced by our perspective, background knowledge, education and social position; they are never neutral. Thus, from this perspective, Kant's thought that beauty is a universal value, anchored in our rational nature, would definitely be questionable.

However, against such a 'postmodernist' critique, the following objections might be raised.

First, to say that the distinction between art and non-art is an ideological *construction* does not imply that the distinction is not valid. In fact, human practices (science, philosophy, religion, art etc.) may fail to have value 'in themselves' and may yet be of value and be respected by all those who take part in those practices.

Second, the idea that the distinction between art and non-art or beauty and kitsch is 'ideological' makes it difficult to explain why there is a fairly stable list of 'classic' works on the basis of which an aesthetic and/or artistic difference in

value has been acknowledged – between, say, novels such as Gustave Flaubert's *Madame Bovary* and Dan Brown's *The Da Vinci Code*.

Third, the postmodernists confuse interests, goals and needs which can be identified independently of the existence of an artistic practice with interests, goals and needs which are *determined by* a practice. Therefore, they (unjustly) reduce art to 'that which a community decides to call art'. They consider 'art' as a descriptive concept. However, 'art' is an *evaluative* concept, which presupposes that there are really values that come about in and through creating and appreciating art (and that are not incidentally called 'artistic values'). The value of art is constituted by the practice of creating and appreciating art and is also constitutive of the concept of art.<sup>1</sup>

Fourth, the terms and distinctions between 'art' and 'non-art', between 'high culture' and 'low culture' and between 'aesthetically valuable' and 'aesthetically failed' may have been fully acknowledged in the West only since the 18th century, but the discussions had been around for centuries and are not limited to our 'authoritarian', modern, European culture. Similar distinctions were discussed by e.g. Plato, Aristotle, Bharata, Confucius, Augustine and Boethius, and many others everywhere around the world. Claiming that they are *merely* an 18th century, European 'invention' is wrong and unjustly undermines their universal significance.<sup>2</sup>

Fifth, (postmodernist) relativism is usually a mere theoretical stance. In practice, even the fiercest defenders of aesthetic relativism demonstrate aesthetic norms and preferences when, say, decorating their house, deciding which clothes to wear or which places to visit on a city trip. A certain kind of inconsistency, which some may label hypocrisy, seems to be involved in their theoretical dismissal of the 'aesthetic ideology'.

Finally, and this may be the most important argument, relativists often defend their views on the basis of an 'anti-authoritarian' feeling of justice. They aim to resist the taste of the 'ruling classes', giving voice to suppressed minority groups. That goal is noble, but are their strategies convincing? For if we need to accept that 'everything is relative', is discussion still possible? If these 'anti-authoritarian' thinkers consider their judgements immune from criticism, they might themselves be called intolerant and authoritarian since they do not accept (universally acceptable) standards on the basis of which aesthetic disputes might be resolved. For, in their view, taste is relative: we cannot criticise the others' appreciations, since 'anything goes'. Thus, some postmodernists actually

---

1. The first three arguments can be found in a slightly different form and applied to literature in Lamarque & Olsen (2002, 441-442).

2. I borrow this line of thinking from Scruton (2009, 64).

force us to accept their judgements uncritically. According to them, a rational discussion about art and aesthetic matters is impossible, as each judgement is equally valid.

Yet, true tolerance with regard to matters of taste is possible only if we are prepared to let others question our own aesthetic appreciations. A proper discussion about the intersubjective validity of our aesthetic appreciations is impossible, unless we acknowledge that there are correct and incorrect aesthetic appreciations, which is precisely what the relativist denies. Furthermore, as we shall see in the following section, a dismissal of relativism does not imply that the value of an artwork as *art* can be reduced or restricted to its aesthetic features, let alone that only beautiful artworks are artistically valuable.

## 5. VALUES OF ART

In this section, we shall discuss non-aesthetic reasons for valuing art and connect these to theories of art that focus on typical reasons we may have to grant human artefacts the status of *art*: representation (*mimesis*), expression of emotions, moral education, knowledge, and existential significance.

A first theory about the value of art says that art is about representation or depiction (*mimesis*). Indeed, one of the most striking features of many works of art is that they represent events, objects, human beings, and so on. We do not merely see paint blots or pixels or a piece of marble, but we see someone or something in the material that the artist deploys. In addition, the landscape painted may of course also stand for something else, e.g. Arcadia. A still-life may represent death or mortality. This observation has led to the idea that something is a work of art only if it represents or depicts something. This theory is known as the *mimesis* theory of art.

That a work of art is first and foremost a depiction of something is a centuries-old view. Plato and Aristotle already defended it, but the rise of non-figurative and abstract art in the 20th century has undermined its plausibility: it is hard to say what it is exactly that an abstract painting by Mondrian, Rothko or Pollock actually represents. Yet this does not render the *mimesis* theory completely obsolete. For many of the greatest artworks in history are genuine depictions. Not all works of art represent things, but many undoubtedly do: this is still ordinarily the case in photography, film, and sculpture. If we take architecture, music and dance into account, matters become more difficult. For whereas we can argue that a still-life painting of apples represents real apples or, say, objects of temptation in a garden of Eden, it is much harder to find out what it is that a work of architecture, e.g. Canterbury cathedral, represents. Although it may be correct to say that it is a house of God, it does not represent a house of God (see Carroll 1999, 25). The



same difficulty holds for musical compositions: what do Mozart's or Beethoven's string quartets represent? For sure, there is certain music, called programmatic music (such as Vivaldi's *Four Seasons* or Mendelssohn's *Midsummer Night's Dream*), that may be said to represent events (the seasons), but that the *mimesis* theory can capture the essence of *all* artworks is clearly wrong.

A second reason for valuing art is because it succeeds in expressing, arousing and/or stimulating affects and emotions. In his famous sonnet 'Shall I compare thee to a Summer's day?' William Shakespeare expresses his admiration for his beloved (and for poetry) in such a rich, profound and subtle way that many of us will be moved on reading it. The same holds for music: we can be deeply moved by Chopin's *Nocturnes*, because (among other things) of the expressive power of the wistful melodies that we hear. That is why we often share our appreciations for works of art in terms such as 'lively colours', 'a sad melody', 'a cheerful scherzo' etc. We then no longer consider (parts of) the artwork as depictions of things in reality, but as the *expression* of feelings and emotions. Leo Tolstoy (in *What is Art?*), Susanne Langer (in *Feeling and Form*), Robin G. Collingwood (in *The Principles of Art*) and many other thinkers defend varieties of the expression theory of art. In spite of their differences, they have in common the idea that the value of art lies (mainly) in the value of its expression of emotions. The value of a work of art is a function of the emotions that it expresses. As the 19th-century poet William Wordsworth famously states in the Preface to the *Lyrical Ballads*, "All good poetry is the spontaneous overflow of powerful feelings." In this way art may *humanise* us: we succeed not merely in approaching the world with human feelings but also in getting to know, estimate, and refine our own emotions and those of others.

How convincing is this second, 'expression' account of the nature and value of art? It is implausible insofar as it claims that a work can be artistically valuable primarily because it expresses and/or arouses affects and emotions. Numerous artworks convey first and foremost ideas, views and thoughts, and they do not *express* emotions nor are they aimed at *arousing* feelings or emotions in us – think of conceptual art, but also of many novels, films and paintings. Another problem is that it is not that easy to provide a satisfactory philosophical account of emotion. Exactly what should be classified as 'emotions' and how can we discriminate between different kinds of affective states, such as, for example, moods (cheerfulness, depression ...) and emotions (anger, sadness ...)? Moreover, it is often claimed that in some works of art, especially those involving fiction, emotions are not directly aroused: perhaps they are imagined or contemplated. Or maybe merely 'quasi-emotions' are evoked, for when we are watching a play or a film we are not moved to direct action and may even remain in a state of tranquil contemplation (see Schopenhauer 2010, 200-204; Sheppard 1987, 18-37; Walton 1978). Although it is plausible to argue that the expressive power of many works

of art intensifies or even determines their value as art, the expression theory clearly faces a number of difficulties. Therefore, a number of philosophers, such as Rudolf Arnheim, Deryck Cooke, and Nelson Goodman, have developed more nuanced versions of the expression theory that focus on the artwork itself and characterise emotional expression in terms of an artwork somehow *symbolising* or *representing* emotion, regardless of the feelings of the artist and the audience (see Neill 2005, 422-23).

A third reason for valuing art originates in the *moral* value certain works of art possess. Aristotle already argued that a successful tragedy should always possess moral value: the hero should be noble in order that the spectators can identify with him and can be moved by his downfall. Aristotle's famous idea of the (moral) purification or purgation (*katharsis*) is intimately connected to the specific value of Greek tragedy. In contemporary aesthetic theories, too, the relation between aesthetic and moral value is a highly debated issue. Several philosophers argue that moral flaws of an artwork impair its artistic value. One of the more extreme examples may be pornography. An intriguing dispute in the philosophy of art concerns the issue whether pornography can be art and, if so, whether and how it is that pornographic art can be valuable as *art* – think of many paintings by Egon Schiele or a film like *L'empire des sens* (1976).

Many philosophers defend the morally educational value of art. The American philosophers Richard Rorty and Martha Nussbaum, for instance, are convinced that art and literature somehow contribute to our moral education. The British philosopher Roger Scruton (2007) claims that (some good) artworks teach us how we ought to behave properly and what we ought to feel when confronted with injustice, the death of a close friend, the birth of a child, war, family conflicts ... The question remains, however, whether we can really become better people by listening to Mozart, reading *Uncle Tom's Cabin*, *The Picture of Dorian Gray* or *Pride and Prejudice*, studying paintings by Marlene Dumas or Luc Tuymans or sculptures by Auguste Rodin. Even though many artworks might teach or show us to feel what we ought to feel and do what we ought to do, it is not clear whether we *will* actually feel and do what we ought to feel and do when the time is right for it in real life situations. Nazis adoring Mozart or Schubert clearly did not. Moreover, we might raise the question who or what determines what we ought to feel or to do.

A fourth reason to value art pertains to the cognitive rewards one may gain from it. Schelling, Solger, Hegel, and several other 19th-century German idealist philosophers defended some version of the cognitive theory. Some of them also maintained the superiority of philosophy regarding this cognitive function. According to Hegel, the knowledge we may gain from artworks is sensuous and, therefore, inferior to the abstract, conceptual knowledge we obtain through philosophy. Is it true that the *content* of the knowledge conveyed by art and

philosophy is the same, namely the World Spirit, as Hegel claims? Today we are reluctant to follow Hegel in believing that all works of art are merely sensuous vehicles of the universal World Spirit, which will be superseded by or sublated (*aufgehoben*) in religion and philosophy. The Hegelian idealist philosophy of art thus considers art an imperfect expression of metaphysical ideas (and of the essence of the *Weltgeist*, which all ideas ultimately express).

Other, non-Hegelian philosophers also argue that the value of art is primarily cognitive. Contrary to Hegel and some of his followers, philosophers such as Wilhelm Dilthey, Hans-Georg Gadamer, Ludwig Wittgenstein, George Santayana, and Roger Scruton acknowledge the cognitive value of art, denying that philosophy is superior to art. For the ideas that artists express are not necessarily metaphysical or historical truths that can be more clearly conveyed by philosophers than by artists. They are the product of the imagination of the artist, bringing about a free play of the spectator's mental powers. The power of art is such that it may well surpass in certain imaginative ways the conceptual rigour typical of philosophical reasoning. Their conclusion is that art definitely possesses cognitive value – it offers valuable insight into the world, other human beings and ourselves – but this value is of a different nature than that of philosophy and science. The knowledge offered by artworks cannot be translated into philosophical or scientific terms without losing much or all of its meaning. An artwork has meaning in the sense that it is meaningful or relevant to you, perhaps in an ineffable sense, because it somehow (at least temporarily) fulfils your life. Artworks may convey much interesting information, but they are not (cognitively) valuable primarily because of this characteristic. There is a specific kind of cognition that, as Wittgenstein says, cannot be conveyed through concepts but can only be *shown*. On this view, artworks might be said to show what is ineffable.

This highlights a fifth reason to appreciate works of art. Apart from their representational, expressive, moral and cognitive value, works of art have *existential* significance. This existential theory claims the following: great works of art can be considered as evocative symbols whose meaning cannot be transferred to another medium: they can show only what they show in the very way they show it. Only in such a way do they really make sense and can they be significant to human beings. The meaning of an artwork has a non-transferable relation to its medium. Its content cannot be transferred to another medium while still remaining equally meaningful or evocative. Artistic meaning can never be adequately paraphrased in discursive terms. If we paraphrase a poem or recount a film or novel, we cannot ultimately do justice to the meaning of the work *as a work of art*. Works of art are not decorative forms providing paraphrasable contents and transient pleasures, but suggestive symbols with existential value. This implies that they not only represent but may also *transform*

human existence. Great works of art, such as Milton's *Paradise Lost*, Sophocles' *Antigone*, David's *Psalms* and Wagner's *Parsifal*, and their unique and subtle combinations of forms, styles and contents, are said to convey deep meanings that evoke a transfiguring experience of the world.

This fifth value of art, its existential significance, may well be the most difficult to grasp and express. Great artworks have made us see the world in a wholly new perspective – think of Homer's *Odyssey*, Shakespeare's *Hamlet*, Goya's *Black Paintings*, Picasso's *Guernica*, the Pantheon in Rome, the Taj Mahal in Agra, etc. Those masterpieces have altered the way we experience reality. One could contend that such works of art put a spell on us, for they reveal and defuse immense and often untameable powers and emotions. Take the view on sexual desire and love developed by Richard Wagner in his great opera *Tristan und Isolde*. Wagner characterises the tremendous and potentially destructive power of erotic love – not merely in the narrative, but also primarily in and through the music itself – by showing how the two young lovers hunger to be united at any cost, and how they are eventually joined in death. Because of its enchanting rendering of deep emotions and important moral insights concerning the nature of erotic passion, longing and love, the opera may be said to console, elevate and ennoble us. By centring on self-sacrifice as the inexorable engine of love, the work possesses a redeeming quality that seems to lift us temporarily above the world of our own mortal condition as embodied individuals (see Scruton 2004).

It is in this sense that one could argue, with Roger Scruton and others, that experiencing great art is akin to religious experience and the sacred. Not in the sense that art necessarily conveys 'spiritual' messages (whatever that may mean), but because it displays energies and emotions that we would not be able to experience otherwise, and because it defuses what we cannot control: the stunning magnitude of the universe, the messiness of life, the finitude of human existence, and so on. From this it does not follow that the meaning of aesthetic appreciation and art is so mysterious that we cannot say anything about it, nor that it inevitably leads to a religious attitude. It does imply, however, that words may fail to express the unique and irreplaceable intensity and impact of a great work of art. One cannot paraphrase the meaning of a profound work of art without *essentially* detracting from the value of directly experiencing it.

Even though none of the discussed theories can be considered to have fully grasped the essence of (the value of) art, each one of them offers thought-provoking insights and rightly insists that art has an indispensable part to play in shaping the human world. Taken together, the discussed views definitely show why it remains worthwhile not only to engage with works of art but also to treasure great works of art and the rewarding experiences they offer.

## 6. CONCLUSION

**B**eauty is an important human value and is widely considered to give meaning to human life. We have examined the problem of the judgement of beauty and the relativity of aesthetic taste and, subsequently, discussed the different reasons we may have for valuing art. But appreciating beauty and art are activities that raise a number of difficult issues. Beauty can occur in many different objects, contexts and events, and appreciating it raises the problem of the objectivity of one's judgement: it is based on a personal experience and feeling of pleasure, but claiming that something or someone is beautiful seems to be more than merely expressing our individual preference: we require that others agree and seem to ascribe an objective quality to a particular object. This led to a discussion of the nature and worth of the judgement of beauty, as well as arguments pro and contra relativism concerning beauty and artistic value more generally. Furthermore, we examined different accounts of the value of art.

Whatever one's exact position in the debate about relativity/universality and subjectivity/objectivity, it has become clear that the issues are complex and cannot be easily and definitively settled in favour of one or other position. In fact, there are many subtly refined viewpoints that are to be situated somehow between subjectivism and objectivism. Furthermore, artistic practices have become so diverse that it is impossible to identify one (primordial) value that *all* artworks possess.

Moreover, novel insights from the relatively new domains of experimental philosophy, empirical aesthetics and neuroaesthetics will undoubtedly raise new questions about the issues discussed above. Whether this novel empirical methodology will undermine the need for and significance of a purely philosophical approach of beauty, aesthetic judgement, the value of art and related topics remains to be seen, but it will undoubtedly stimulate philosophers to refine their conceptual analyses and revise some of their arguments and hypotheses. The philosophical enquiry into the nature of beauty, art and aesthetic value will thus remain a very lively and important research area, not in the least because the need for a philosophical reflection on beauty and art is crucial to all of us who consider them of paramount importance to a humane and civilised existence.<sup>3</sup>

---

3. I wish to thank Bart Engelen, Martin van Hees, Lodi Nauta and Veerle Rotsaert for their comments.

## RECOMMENDED FURTHER READING

- Bourdieu, P. (2010). *Distinction: A social critique of the judgement of taste*. Routledge. (Original work published 1979)
- Carroll, N. (1999). *Philosophy of art: A contemporary introduction* (A. C. Routledge. Danto, Ed.). Open Court.
- Eagleton, T. (1990). *The ideology of the aesthetic*. Blackwell.
- Guyer, P. (2014). *A history of modern aesthetics. Vol I-III*. Cambridge University Press.
- Kant, I. (2000). *The critique of the power of judgment*. Cambridge University Press. (Original work published 1790)
- Kieran, M. (2005). *Revealing art*. Routledge. <https://doi.org/10.4324/9780203642351>
- Lamarque, P., & Olsen, S. (2002). Truth, fiction, and literature. In J. Levinson (Ed.), *The oxford handbook of aesthetics* (pp. 421–435). Oxford University Press.
- Santayana, G. (1955). *The sense of beauty*. Charles Scribner's Sons/Macmillan Publishing Co. <https://doi.org/10.1037/10861-000> (Original work published 1896)
- Schopenhauer, A. (2010). *The world as will and representation* (Vol. 1). Cambridge University Press. (Original work published 1818)
- Scruton, R. (2004). *Death-devoted heart: Sex and the sacred in Wagner's Tristan and Isolde*. Oxford University Press.
- Scruton, R. (2007). *Culture counts: Faith and feeling in a world besieged*. Encounter Books.
- Scruton, R. (2009). *Beauty*. Oxford University Press.
- Sheppard, A. (1987). *Aesthetics: An introduction to the philosophy of art*. Oxford University Press.
- Walton, K. (1978). Fearing fictions. *The Journal of Philosophy*, 75(1), 5. <https://doi.org/10.2307/2025831>

MARTHA NUSSBAUM



## GEORG W.F. HEGEL

### VORLESUNGEN ÜBER DIE PHILOSOPHIE DER KUNST

---

Hegel's series of lectures on the philosophy of art has been a landmark in the history of aesthetics ever since its publication in 1835 and has had an immense impact on a wide variety of thinkers, including Rosenkranz, Vischer, Dewey, Heidegger, Croce, Adorno, Lukács and Danto. Hegel is a metaphysical idealist, who claims not merely that artworks express divine and human freedom and ultimately embody the so-called 'absolute spirit' or the ultimate rationality of reality, but also that the material aspects of art are ultimately dispensable for thoughts that can be expressed more appropriately through the more intellectual means of religion and philosophy. Fine or 'beautiful' art (*schöne Kunst*) is defined as 'the sensible shining of the idea'. Art, religion and philosophy have the same content (i.e. the freedom of the spirit), but different forms. Despite its grand metaphysical claims and its clearly cognitivist approach, Hegel's philosophy of art provides often stunningly detailed studies of specific artworks, e.g. of Flemish 15<sup>th</sup>-century paintings by Memling and the Van Eyck brothers. He also offers an historical account of the value of art that formed the basis of the 19<sup>th</sup>-century discipline of art history. Hegel distinguishes three art forms: the symbolical (e.g. Egyptian), classical (e.g. Ancient Greek) and romantic (e.g. mediaeval) art forms. The latter art form, which Hegel mainly situates in Christian painting, announces the notorious 'end of art'. The outward forms of the artwork have become superfluous. Hegel argues that for us art has become something of the past and has to be superseded by religion and philosophy, for (in his view) any material, sensible medium is inadequate to capture fully the ultimately spiritual essence of reality.

---



Corijn van Mazijk

## **Examining Life**

# EXAMINING LIFE

## *An Introduction to Phenomenology and Existentialism*

Corijn van Mazijk  
[c.m.a.van.mazijk@rug.nl](mailto:c.m.a.van.mazijk@rug.nl)

### 1. INTRODUCTION

**W**hat does it mean to exist? How do we live our lives meaningfully, if life has a meaning at all? In philosophy, these kinds of questions are called 'existential questions'. They are the sorts of questions human beings may be inclined to ask when existence itself becomes a problem for them, while the answers provided by tradition have lost their persuasiveness. Existentialism is a subdiscipline of philosophy that reflects on such questions. In the 20<sup>th</sup> century, existentialism rose to prominence particularly in dialogue with another form of philosophy, namely phenomenology: the then prevalent form of philosophy in Germany, founded by philosopher and mathematician Edmund Husserl (1859-1938).

This close kinship of phenomenology and existentialism can hardly be called a coincidence. Phenomenology is, broadly conceived, the study of subjective experience. It is this subjective experience that the existentialists took as their starting point for considering the nature of human existence, and to address some of the urgent philosophical concerns that are connected to our existential situation, in particular our finitude, the meaning of life, and the problem of freedom.

To understand existentialism in its various shapes, then, requires a study of its relationship to phenomenology. Starting from Husserl, the father of phenomenology, this chapter introduces the reader to some of the most important themes discussed in both phenomenology and existentialism.

## 2. THE PHENOMENOLOGICAL PROJECT

If you take a look around you, you will see all sorts of things: perhaps a desk, a coffee mug, your computer, and so on. All these things present themselves to you as being real. That is, in ordinary perception, you seem to have immediate access to those real things out there. You can see them, feel them, smell them – there is hardly a good reason, at least under usual circumstances, to doubt that these things are really there. The totality of the things that we thus perceive constitutes what we normally call ‘reality’.

Reality in the sense of the world of real things around us has been the object of scientific and philosophical study for centuries. Put simply, biology studies organisms, physics laws of matter and motion, neuroscience the brain, and so on. According to Husserl, all these sciences overlook something important – even more important than what their own respective domains contain – namely subjective experience. Consider again the objects that surround you. Indeed, you see them; they present themselves to you as real. But reflection shows that this ‘reality’ is ultimately nothing more than a ‘sense of reality’. You don’t access these real things from outside of your experience; there are incessantly given *through* it. Things therefore ‘present’ themselves to you as real in your experience; their ‘reality’ turns out to be nothing more than a *type of presenting* in your experience. We could call this ‘presenting as real’, and contrast it to ‘presenting as not real’, which then applies, for instance, to freely imagined objects, or hallucinatory objects.

Like many German philosophers of the century before him, Husserl took this thought seriously. In phenomenology, there can be no object without subject. In other words, phenomenologists believe that reality must be understood from within the order of the conscious experiences of subjects. To ask what the world would be like independently of our experiences, is regarded as futile and meaningless.

Husserl even took this insight one step further. According to him, science and philosophy before him, as well as all our everyday behaviour, take place in what he calls the ‘natural attitude’. This term refers to our ordinary way of experiencing the world, which is characterized by the naive acceptance of an independent reality of things. Put differently, our ordinary experience so expertly erases its own traces in presenting us things outside of us, that it makes itself invisible in

this process. In other words, we forget that our experience accomplishes our ordinary sense of reality. According to Husserl, this trick is so skillfully performed that it has taken Western philosophy thousands of years to finally step out of it, and to see that reality is actually something 'constituted' in experience.

The phenomenological method commences with the attempt to systematically outline a way for an individual to take in an alternative position to the natural attitude. Generally speaking, this method is called 'phenomenological reduction'. Husserl's ambition is to develop a science of our experiences, phenomenology, that is more radical than the regular sciences since it breaks free from the natural attitude. The phenomenological method that is to accomplish this new science involves three steps.

First, (i) the phenomenologist has to be truly honest. He or she has to describe the way in which things unfold in experience (whether natural things in the external world, imagined objects, mathematical objects, and so on – whatever passes before the mind) without reference to any convictions that do not form a part of the experience itself, such as scientific theories, religious beliefs, and also the belief in the existence of the world in general. This step is called 'epoché', a kind of bracketing or suspending of beliefs and theories which are extrinsic to the experience. The epoché is a radical procedure that cannot be learned overnight; it involves suspending everything one has learned from science, for instance, regarding the impact that natural objects have on the human body, and even the belief in the existence of the mind-independent world. After the epoché, one would be left with nothing but 'pure experience' (which still includes the objects experienced, but now purely as experienced).

Second, (ii) the phenomenologist would have to find the right words to describe what is thus observed. This is particularly difficult and problematic, since most of our language derives from the natural attitude. For instance, when we describe the colour red that is at the end of the visible spectrum of light, we do not describe it as we experience it purely when looking at a red object. To do so, the phenomenologist must invent a new vocabulary to describe the experience. While this is a challenging task, Husserl and his colleagues did make serious headway in the creation of a new terminology.

Thirdly, (iii) the phenomenologist should not just describe whatever happens to pass by in his or her mind, but instead try to universalize statements about this. For instance, I won't describe how a desk appears to me in perception, but instead how *any perceptual object in general* is bound to appear to anyone in perception. Put differently, while I might start from just any particular perception (or imagination, calculation, etc., depending on the type of mental act I'm studying), I will try my best to generalize my claims so as to make them valid for a whole class of mental acts. This way, claims about 'subjectivity' can gain 'objectivity',

in the sense of a universal character that makes them valid for everyone. The operation by which we thus arrive at the essence of a phenomenon, as for instance the essence 'perceptual object in general', is what Husserl calls 'eidetic reduction'.

More generally, most of the phenomenologists after Husserl are skeptical about the possibility of performing any of these steps perfectly. For instance, Heidegger (in *Being and Time*) and Merleau-Ponty (in *Phenomenology of Perception*) performed some kind of epoché as well as eidetic reduction. That is to say, first, that they did engage in the practice of some form of phenomenology through an epoché. Like Husserl, they are critical of the natural attitude, and seek instead to describe simply how things appear to us. Second, they seem to perform something akin to an eidetic reduction as well, as they often make claims that pertain to human existence in general, rather than just to their own personal existential situation. However, unlike Husserl, they dispute the possibility of performing either reduction perfectly. For them, no description of experience is ever completely pure and unbiased.

For phenomenologists, all objects are ultimately experienceable objects, such that there is no Kantian 'thing in itself'. For this reason, Husserl thought that the phenomenological study of experience was the only true 'first philosophy'. Since the world studied in science is only given through experience, the study of experience – phenomenology – has absolute priority over all the other sciences. Moreover, all philosophical problems could be resolved simply by accurately describing how things unfold for us in experience. All theory and argumentation is just a distraction from the pure experiences.

In focusing on human experience, the phenomenological approach in philosophy led many subsequent philosophers to consider human existence in more detail. In the next sections, we will look at some aspects of this existentialist turn in philosophy, and how they shift focus to concerns over mortality, the meaning of life, and freedom.

### 3. EXISTENCE AND FINITUDE

**L**et's start again by taking a look around. Suppose once more that among other things, you see a desk, a coffee mug, and your computer. Now that I've brought your attention to these things, they appear to you as standing over-against you. It seems that they will remain to be there independently of whether you're actually there to see them or not. They stand before you dryly, silently, and unmoving, as the collection of inanimate objects they are. However, just before you had turned your attention to them, these things were of course already present to you, albeit in a very different way. Before, you did not consider

them explicitly: you may have been *occupied* with them – occasionally drinking coffee from the mug, and checking Facebook on your computer. You were *using* these things, absorbed in your practical activities – but strictly speaking you were not considering them as the things which they have become now that your attention has shifted to them.

Following Heidegger's influential existentialist phenomenology in *Being and Time* (1927), things have two modes of being: 'present-at-hand' and 'ready-to-hand'. That is to say, they either appear as objective things (roughly speaking, as in Husserl's natural attitude), or they appear in practically meaningful relations to us. The thing in its practical way of appearing (the ready-to-hand) is also called 'tool'. In this sense of the word, anything can become a tool by appearing as ready-to-hand. This distinction is to some extent comparable to what contemporary philosophers sometimes call 'knowing-how' versus 'knowing-that'. While you know *that* the coffee mug is an object used for drinking coffee from, you also know *how* to use the mug in order to practically facilitate the drinking of coffee. The first type of knowledge could also be called theoretical knowledge, the second skillful knowledge. Heidegger's phenomenology gets its first existential impetus by arguing that human beings are always already involved in practices. This means that, in waking life, you're not just continuously using things, but those things in their being used refer to your ongoing actions and preferences as well as to other things ready-to-hand. This constitutes a so-called 'context of significance'. For instance, before anything becomes present-at-hand to you, the coffee mug presents itself as a drinking tool. But the mug as tool-object also bears a relation to the coffee machine – again not an objective (present-at-hand) relation, but as a practical relation. You, the coffee mug, and the coffee machine can be said to belong to one holistic *context of significance*, and importantly, they do so before we start thinking about or reflecting upon coffee, mugs, or machines.

According to Heidegger, the existing human being (by him famously called *Dasein*) always finds itself already involved in some context of significance, whether we realize it or not. This further means that, whenever we interpret the world objectively (as present-at-hand), be it in sociology or in theoretical physics, we do so from within a certain practice. Indeed, all our interpretations are inevitably colored by the practical circumstance we're in. This constitutes a clear *hermeneutical* element of Heidegger's phenomenology, which isn't found in Husserl's: namely that all human knowledge is a practically constrained *interpretation*.

Heidegger claims that this involvement in practices which defines the human standpoint is inevitably shaped in various ways by time (hence the word 'time' in his main work *Being and Time*). First of all, our past plays an important role in determining our existential situation. For Heidegger, in contrast to Husserl, it is

mainly the past as *history* that determines human existence. For instance, I was born in the late twentieth century. Had I been born in the 16<sup>th</sup> century, my thoughts, actions – indeed my whole existence – would have been fundamentally different. In this way, the human being is, prior to any choice on the matter, ‘thrown’ into history (to use Heidegger’s term), just as it finds itself ‘thrown’ into a particular practical setting each time it awakens to reflect upon things.

The future, however, also determines human existence. In the second division of *Being and Time*, Heidegger argues that our understanding of ourselves and the world is always done from the finite standpoint of mortal human beings. Human beings know that they will die. This is not just a negative feature of existence; it also makes the time that we do have meaningful. It is, for instance, a great joy to meet with old friends, but part of what makes this so special is that we won’t be able to keep seeing our friends time and again for all eternity.

Many philosophers have tried to gain insight into eternal truths; to transcend the finite human standpoint in order to achieve eternal knowledge, whatever they thought that would be. With its existential twist, that is to say, by focusing on the human being in the here and now, Heidegger’s phenomenology brings all knowledge back into concrete, finite, human existence. According to Heidegger, any understanding of the world is an interpretation from the standpoint of a finite temporality; of a subject engaged in a practice, thrown into history, and facing an inevitable death.

Heidegger’s phenomenology receives its second existential impetus from his account of the average human being’s existence and the possibility of authenticity. That is to say, while most of his philosophy outlines the universal human (practical) standpoint from which we encounter the world and each other, Heidegger divides human beings into those who occupy an inauthentic mode of existence and those who occupy an authentic one. Because of their social nature, human beings tend in their average everydayness to unreflectively take over the opinions of others. Whether in our taste of music or our political preferences, we tend to safely follow others rather than to truly develop our own perspective. In particular, this brings us to cover up the inevitable promise of our own death. *Dasein* flees from its own death; following the viewpoint of the masses, it pretends that death is something that won’t matter until the very end, or it acts as if death only concerns others, but not oneself.

For Heidegger, the fearful facing of one’s own being-toward-death allows one to transcend this flight as an inauthentic mode of being, and to take back one’s own life from the dominating opinions and superficial small-talk of the masses. Because my death is an inevitable possibility that I cannot avoid, and that nobody can take up in my place, facing it at the cost of fear can be a revelatory, individualizing experience. Certainly, it won’t change the fact that my existence is,

as he famously calls it, 'thrown' and my knowledge fallible; I never will transcend the mortal standpoint. I will, in the end, continue to act in ways prescribed by my historical time and by others by whom I am surrounded. Furthermore, I will continue to inevitably find myself dwelling in one practical situation or another. Nevertheless, by taking back my own individual life by facing my inevitable death, my existence would gain a certain character of 'resoluteness', which would be typical of an authentic way of being. Perhaps more than anything else, this account of authenticity has shaped the philosophical subdiscipline of existentialism.

#### 4. EXISTENCE PRECEDES ESSENCE

For the French existentialist Jean-Paul Sartre (1905-1980), all the things that you currently see are there in a manner which he calls 'being in-itself'. By this, he refers primarily to the ordinary objective existence of material (non-conscious) things around us, just as Heidegger had spoken of the present-at-hand. In his main work *Being and Nothingness* (1943), this is contrasted with so-called 'being for-itself', which would be typical of human consciousness: the being which is given to itself. Sartre's ontology is thus based on a simple distinction between non-conscious and conscious (in the sense of humans) beings.

In an earlier essay called 'The Transcendence of the Ego' (1936), Sartre had already prepared the way for this division. Here, he argued that Husserl was wrong to think that experience or consciousness must involve something like a subject or ego. Instead, if we follow Sartre, consciousness relates to all sorts of things (desks, coffee mugs, computers), but it is itself wholly empty. This can be illustrated in the following way. On Sartre's view, if you reflect upon your own consciousness, you will find nothing but just the *things* outside of you that you are conscious of; consciousness 'itself', on the other hand, is 'nothing'.

It is this view on the nature of consciousness that motivates Sartre's later theory of freedom. To be sure, human beings are in some sense material objects; they have physical bodies, and they constantly act in the physical world. Existentialists call this our 'facticity': the totality of concrete, unchosen facts that define one's individual existence. However, our facticity notwithstanding, Sartre's existentialism continues to ascribe a radical freedom to human existence. This freedom is radical in the sense that we cannot escape it (we are 'condemned' to it), but also in the sense that it is unlimited. This is best explained by a few examples.

One of Sartre's famous examples to illustrate existential freedom is of a waiter working in a restaurant. As is clear from his particular way of speaking and



gestures, the waiter is *acting* as a waiter. While the setting of a fancy restaurant may rightly motivate this person to take up a certain role, the waiter in this example loses himself in this role; he is trying too hard to be a waiter, and is thereby pretending to be something he is not. In a way, he could be said to pretend to be an object: namely a waiter, something with determinate qualities, styles of behaving, etc. In reality, however, according to Sartre at least, he is precisely not an object (a being in-itself), but a being-for-itself, and therefore he is free to determine himself. In other words: by pretending to be a 'thing' (waiter), this person would avoid the freedom to choose for himself who or what he really is. The waiter therefore acts in what Sartre calls 'bad faith' – his version of Heideggerian inauthenticity.

For Sartre, the example of the waiter should serve as an illustration of something many of us do all the time. Because our freedom to determine ourselves can be scary to face, we tend to lose ourselves in all sorts of roles prescribed to us by other people or by society – whether the role of a son, a mother, an academic, or something else. We make life easy for ourselves by playing out these roles, thereby falsely turning ourselves into things. For humans, however, 'existence precedes essence'. While we can try to act out life in bad faith, making ourselves believe that we are things with clear 'essences', human beings are in the end condemned to the freedom to determine their own being.

For Sartre, in spite of the fact that motives may tempt one to act in one way or another, there always remains in the end a freedom 'between' motive and response. While it is true that we are determined by our past and present circumstances, human beings are still condemned to this freedom. A former gambling addict, when confronted with the opportunity to gamble in a casino, is again confronted with this freedom. The decision to stop gambling needs to be made anew time and again, because there's always the free choice of starting again. A final, and similar, example: what's scary about standing at the edge of a high building? On Sartre's account, it's not the distance that you fear – it's the fact that you are always free to really make the jump.

## 5. FREEDOM AND THE LIVED BODY

**M**erleau-Ponty's (1908-1961) existentialist account, likewise originating in dialogue with phenomenology, takes us into a very different direction. As we have seen, Sartre distinguishes human beings (the for-itself) from the world they live in (the in-itself). Merleau-Ponty, much more like Heidegger, is keen to avoid such a duality. However, instead of starting, as Heidegger did, from the practically engaged human being, Merleau-Ponty starts from the 'lived body': your own body as experienced by yourself.

This lived body might be said to replace what Husserl called pure consciousness and what Heidegger called *Dasein*, which formed their respective starting points of analysis. Following Husserl, Merleau-Ponty distinguishes the lived body from the objective body. While the latter is the body as an object (as considered in the natural attitude, or as present-at-hand), the former is the body as we live it; it is our point of contact with the world. For Merleau-Ponty as for Sartre, there is no little homunculus (little human being) locked away somewhere in a special place called the mind, from where it observes everything that happens. Instead, consciousness *includes* the lived body, and to a large degree *is* the lived body in its interactions with the world.

Even more so than Husserl and Heidegger, Merleau-Ponty emphasizes the contact between our bodies and the world is shaped by habits, skill, and practices. This feeds into his existentialist account of human freedom. For Sartre, human beings were free in the sense that they can always transcend any factual situation, so that they are condemned to free choice. According to Sartre's existentialism, there are obstacles to our freedom only where we have freely interpreted things as being obstacles for us. This is because, on his view, we are always free to give meaning to our own lives.

Merleau-Ponty's existentialist views on freedom are different. According to him, Sartre's conception of radical freedom is too abstract. As a consequence, it would not sufficiently account for the many ways in which our bodies constrain our freedom. For example, if a man is drowning, a bystander with no swimming experience will not feel the same degree of freedom and ability to save that person as an expert swimmer would. In the same way, someone raised to abide by the law and the customs of society will be much more reluctant, for instance, to steal something from the supermarket than someone who has not been thus raised. Finally, someone who is in great bodily pain can no longer, at some point, transcend that situation to again experience joy in life.

Sartre's point, that there is always freedom 'between' the motive and the act, doesn't seem to do justice to these observations. Perhaps the poor swimmer and the good citizen can entertain the options of saving the drowning man and stealing from the supermarket respectively. They can transcend their factual situations, and entertain all sorts of options in free thought. But, in the end, what good does this type of freedom do them? Like Heidegger, then, Merleau-Ponty reminds us of the finite nature of our existence, not just because of our historical situation, but also because of the limitations of our own individual bodies.

## 6. THE EXPERIENCE OF THE OTHER

For the last time, let us consider the things experienced in our immediate environment. Perhaps, this time, there's not just a desk, a coffee mug, and computer, but also a person sitting over against you. How does the experience of another person differ from the experience of mere things? According to the existentialist Emmanuel Levinas, the difference is what defines us as human beings.

The encounter with the other is one of the main themes of *Totality and Infinity* (1961), in which Levinas develops an existentialist ethical theory. According to Levinas, experiencing the face of the other is fundamentally different from all other human experiences. Ordinary things are objects for me, but I am no object for them. While I might do with objects as I please, the face of the other is neither controllable nor predictable. Moreover, the other is never wholly present to me; I cannot take in the other's point of view. The thoughts, feelings, indeed the whole existence of the other is something I cannot genuinely gain access to; I am, after all, stuck to my own point of view. Therefore, the other is not something that belongs to my material world in the way ordinary objects do. Instead, in the face of the other, I encounter what Levinas calls 'infinity'.

According to Levinas, the other's gaze also makes a new kind of appeal onto me; it places a certain responsibility upon me that is absent in my interaction with ordinary things. Moreover, according to Levinas the subject only truly becomes subject thanks to the encounter with the other. This encounter is the starting point of our existence, which would otherwise be a mere dwelling in a world of things. It is also the basis of problems of human freedom and responsibility. Even the possibility of objective knowledge rests on the encounter with the other. After all, without others, there would be no shared norms by which to assess things; everything would be a matter of my own perspective.

For Levinas, the phenomenological description of the encounter with the other does not simply result in a particular ethical theory like utilitarianism, deontology, or virtue ethics. It is rather a meta-ethical theory that describes the conditions of any ethical reflection. Indeed, for Levinas, the description of the encounter with the other even amounts to a metaphysics, since for him the experience of infinity and the ethical appeal of the other come prior to any questions about knowledge, truth, or reality. After all, we only become questioning, individual human beings through the encounter with the other.

## 7. CONCLUSION

**A**s we have discovered in this chapter, the ordinary things around us take on a new dimension in the phenomenological and existentialist philosophies we have considered. In phenomenology, only the subjective experiences of these or any other objects matter. Whereas sciences leap over the experience to focus solely on the objective thing taken as a real thing – something which phenomenology considers only as a type of experience –, phenomenology sticks entirely to the phenomena as experienced, whether of real things or otherwise. Thereby a whole new field of scientific inquiry is opened up, that aims at describing the general ways in which things appear to us.

Starting with Husserl, phenomenology became one of the central movements of 20<sup>th</sup> century Western philosophy, and it remains a dominant force in contemporary philosophy. Today, phenomenology finds a wide range of applications within the humanities, such as in the field of gender studies, but also within psychology, psychiatry, neuroscience, and artificial intelligence.

To a large extent, existentialism grew out of a confrontation with phenomenology. The most important contributions to existentialism highlighted in this chapter were developed in response to Husserl. While existentialism is generally critical of the bigger ambitions of the phenomenological project, some form of phenomenological method remains central to the existentialist tradition. By using this method to reflect upon the human condition, existentialism has proven itself a branch of philosophy that offers thought-provoking reflections on the meaning of life, while also making significant contributions to metaphysics, ontology, and ethics.

## BIBLIOGRAPHY

- Heidegger, M. (2012). *Being and time* (E. Robinson, Ed.; J. Macquarrie, Trans.). Blackwell Publishing. (Original work published 1927)
- Husserl, E. (1960). *Cartesian Meditations: An Introduction to Phenomenology* (D. Cairns, Trans.). Martinus Nijhoff. [https://doi.org/10.1007/978-94-017-4952-7\\_1](https://doi.org/10.1007/978-94-017-4952-7_1) (Original work published 1931)
- Merleau-Ponty, M. (2012). *Phenomenology of Perception* (D. A. Landes, Trans.). Routledge. (Original work published 1945)
- Sartre, J. P. (1991). *The transcendence of the ego: An existentialist theory of consciousness* (R. Kirkpatrick & F. Williams, Trans.). Hill and Wang. (Original work published 1937)
- Sartre, J. P. (2003). *Being and nothingness: An essay on phenomenological ontology* (H. Barnes, Trans.). Routledge. (Original work published 1943)

EDMUND HUSSERL



HUSSERL, E. (1983)

*IDEAS PERTAINING TO A PURE PHENOMENOLOGY AND TO A  
PHENOMENOLOGICAL PHILOSOPHY*

---

Edmund Husserl's *Ideas I* (full title: *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book: General Introduction to a Pure Phenomenology*) offers the most comprehensive introduction to the philosophical science called phenomenology, and as such it is one of the key texts of twentieth century philosophy. In it, Husserl explains that phenomenology is the science of phenomena: the study of whatever is immediately given to one's own consciousness, regardless of whether the phenomena in question exist or not. Phenomenology's scientific character is, as Husserl explains, based on it being a science of essences: a science of the invariable ways in which differently classified phenomena must appear to any consciousness. In addition to this, *Ideas I* marked Husserl's public turn to a transcendental philosophy. For Husserl, this meant that phenomenology would have the capacity to offer a philosophical foundation for the other sciences, thereby giving phenomenology a certain priority over them. Although many of Husserl's followers were reluctant to accept this transcendental turn, *Ideas I* exerted an incredible influence on twentieth century philosophy, and many well-known philosophers developed their ideas in dialogue with *Ideas I*.

---

---

# THINKING ABOUT THINKING

---



Tim De Mey

**The Quest for  
Truth**

# THE QUEST FOR TRUTH

## *An Introduction to Epistemology*

Tim De Mey  
[demey@esphil.eur.nl](mailto:demey@esphil.eur.nl)

### 1. INTRODUCTION

Suppose, for the sake of the argument, that the White House is occupied by a bully president, whose blunt lies not only ever increase in number but also in preposterousness. Or suppose that, to subvert the call for political or legislative action, “merchants of doubt” call into doubt well-established scientific facts such as climate change or evolution. Or suppose that almost every dramatic, politically sensitive event triggers a proliferation of conspiracy theories. Or suppose that journalists and the media continuously get bad press because they are accused of producing fake news. Or still, suppose that although in theory, communication technology allows for maximum freedom of expression of opinion in an ideal marketplace of ideas, in practice, the real marketplace of ideas blocks rather than facilitates the free and open exchange of views.

Our handful of examples illustrates that one needs little imagination to conceive of real-world situations in which finding out what is true, and what isn’t, is of pivotal, societal importance. Can philosophers be of any help in this quest for truth? Philosophers have always been mesmerized by truth. Part of their interest in truth is focused not on trying to find out what is true and what isn’t, but rather on the end-products of such inquiries, i.e., knowledge, to know what is true and what isn’t. In this context, philosophers have developed and defended different

theories of knowledge, i.e., rival views on the nature, value, possibility, structure, sources, and kinds of knowledge.

The subdomain of philosophy that is concerned with the theory of knowledge is called epistemology, from the Greek word *episteme*, knowledge. It's one of the two pillars of theoretical philosophy, the other being metaphysics (see the chapter on ontology). Other subdomains in theoretical philosophy, like, e.g., philosophy of science (see the chapter on philosophy of science) or philosophy of mind (see the chapter on philosophy of mind), raise epistemological and metaphysical questions in more specific contexts like science and the mind. One might even argue that in practical philosophy as well, the most fundamental philosophical questions are, deep down, epistemological and metaphysical questions. For example, when we evaluate someone's actions morally, their intentions often play an important role, but can we ever know someone's true intentions? And what about moral responsibility? Is such a thing even possible when free will and causal determinism seem to be incompatible with one another?

In this chapter, we will consider some big questions in epistemology. In Sections 1 and 2, we will discuss answers to the following questions: What is knowledge and is knowledge possible at all? Subsequently, in Sections 3 and 4, we shall look at opposing accounts of justification. Reliabilism, the account of justification discussed in Section 4, shifts our attention away from accounts of what knowledge is to the methods by which we acquire it and that most often result in true beliefs. In the last section, we shall suggest that if epistemology aspires to be of some fundamental help in solving pressing real-world problems, like the ones mentioned above, it should focus more on the context of inquiry by unraveling and pinpointing "heuristics," or problem-solving strategies, and by aligning these novel findings with the earlier results from traditional epistemology.

## 2. RIGHT FOR THE WRONG REASONS

One of the central problems in epistemology is the analysis of knowledge, i.e., the question "What is knowledge?" According to the traditional analysis, which has already been considered in Plato's *Theaetetus*, the epistemic subject *S* knows the proposition that *p* if and only if (1) *S* believes that *p*, (2) *p* is true or, in other words, it is a fact that *p*, and (3) *S* is justified in believing that *p*. For instance, you *know* that you are reading this chapter because you experience reading this chapter. That experience, in turn, justifies your belief that you are reading this chapter, and your belief is also true: you are in fact reading this chapter. This traditional analysis has also been called the tripartite or the JTB-analysis, i.e., knowledge amounts to Justified True Belief.

In the traditional analysis, the function of the justification condition is to rule out epistemic luck, i.e., cases in which it's only accidental, coincidental, or fortuitous that the belief of the epistemic subject happens to be true. Consider, e.g., the fictional, lying president hinted at in the introduction (any resemblance to actual presidents is of course purely coincidental). Suppose that out of 1000 of his statements, 999 are lies and only one is actually true. Suppose, furthermore, that an obstinate, gullible adherent believes every claim the president is making, and so, not only believes the 999 falsehoods, but also the president's rare, true claim. Doesn't it seem counterintuitive to conceive of that accidentally true belief as knowledge? According to the JTB analysis, it does, and the reason is that the gullible adherent lacks proper justification.

However, is the mere justification condition enough to rule out all conceivable cases of epistemic luck? In his famous three-page paper "Is Justified True Belief Knowledge?", Gettier (1963) questions whether the three conditions of the traditional analysis are jointly sufficient for knowledge with two quite specific counterexamples. Here we'll interpret an earlier counterexample by Russell (1948) along the same lines: Russell's stopped clock. Suppose that a man looks at an analogue clock that has in fact stopped, though he thinks it is working, and looks at it exactly at the moment when it tells the correct time. All three conditions of the traditional analysis of knowledge have been met: the man acquires a justified true belief as to the time of day. However, it's counterintuitive to conceive of the man's justified true belief as a genuine case of knowledge. In other words, there are cases of epistemic luck that are not ruled out by the justification condition as such.

Attempted solutions to this Gettier problem try to explain why subjects in Gettier cases do not know that  $p$ , despite having a justified true belief that  $p$ . Some add a fourth condition to the JTB-analysis, i.e., an X-factor that marks the difference between genuine cases of knowledge and Gettier cases. One attempted solution is the "no false lemmas analysis": on top of justified true belief, knowledge requires that the belief may not be inferred from any falsehood. The no false lemmas approach solves Russell's clock: the man doesn't know the time of the day because, while the clock has stopped, he erroneously assumes it is still going. Another attempted solution is the "no defeaters analysis": on top of justified true belief, knowledge requires that there may not be any undermining evidence unavailable to the epistemic subject. In the case of Russell's clock, there is such a defeater: if the man knows the clock has stopped, he wouldn't acquire his justified true belief as to the time of the day.

Another strategy to solve the Gettier problem seeks to replace the justification condition. For instance, Goldman (1967) suggests that instead of  $S$  having *justification* for her belief that  $p$ , the causal condition must be met; i.e., the belief that  $p$  must be *caused* by the fact that  $p$ . Goldman's causal analysis solves

Russell's clock: although the man's justified true belief is caused by the time indicated by the clock, the indicated time is not caused by the actual time of day, but by the fact that the clock has stopped. Replacing the justification condition is also the strategy that Dretske (1971), Goldman (1976), and Nozick (1981) adopt, proposing the sensitivity condition instead, which requires that S would not believe that  $p$  if  $p$  were false. Truth sensitivity solves Russell's clock as well because if the actual time of day differed from the time indicated by the clock, the man would still believe the indicated time to be right.

Although all the alternative analyses mentioned above can successfully deal with Gettier cases like Russell's clock, none of them have proven immune to counterexamples up until now. According to Williamson (2000), this predicament results from attempting to add an objective condition, i.e., the truth condition, to the analysis of knowledge, which is rather entirely subjective. Another trend in epistemology pointing in a similar subjectivist direction is standpoint epistemology. For example, Sandra Harding (1991) draws our attention to the social group to which the epistemic subject belongs, and whether that group is marginalized.

### 3. SKEPTICAL ALTERNATIVES

**A** second central problem in epistemology is whether knowledge is attainable in the first place. Meeting the belief condition is not hard at all, but meeting the truth condition may be impossible. Can you know that it's true that the external world and other minds exist? It's conceivable and thereby at least logically possible that you are the only mind that really exists, and that other minds and the external world are merely your confabulations. Can we really exclude such a skeptical alternative? Can we know for sure that it doesn't obtain? And if we cannot rule out such skeptical alternatives, are there any basic things we can establish as certainties, like that we have hands or are reading this chapter?

In medieval times, skeptical arguments were offered to discuss the notions of knowledge and certainty (Adriaenssen 2013). But it is Descartes who is credited with having defined the problem of skepticism with skeptical alternatives. In his meditations, starting from the well-known experience of waking up from a dream and realizing it was all a dream and the subsequent question of how we know that we are not dreaming the whole time, Descartes asks us to take one step further and imagine that we are systematically misled by an evil demon. Whereas the thought that we might be dreaming threatens only the existence and our knowledge of the external world, such an evil demon would undercut even something as sure as mathematical knowledge. The logical possibility of the dream and the evil demon constitutes skeptical alternatives, i.e., alternative

explanations of our experiences that endanger whatever we think we know. While we need to rule them out, we can never do so completely. Recall, e.g., the sensitivity condition: you cannot know that you are not systematically misled by an evil demon because if you were, you would still believe that there was no evil demon.

Let's develop a similar but more recent example. Suppose that you are the epistemic subject *S* and that the proposition *p* that you know is that you have hands. Enters the skeptical alternative. It's conceivable and thereby at least logically possible that you are not a person of flesh and bones, but rather a Brain In a Vat (or BIV) in which all the experiences that you have are artificially produced by electric stimulation. Of course, you do not really believe that you are a BIV, but you cannot exclude the possibility either. So far so good; it's possible that you know that you have hands and that you do not know that you are not a BIV at the same time. However, the skeptical challenge arises from the conditional assumption that "if you know that you have hands, then you know that you are not a BIV." The underlying idea is that knowledge is factive: you can only know that you have hands if it's a fact or if it's true that you have hands. In other words, you can only know that you have hands if it's true that you are not a BIV, but rather a person of flesh and bones (and hands). So, if you really know that you have hands, then it is *implied* that you know, or at least are in a position to know, that you are not a BIV.

The argument of the skeptic is that being able to know that you are not a BIV is required for you to know that you have hands, and since you cannot know that you are not a BIV, you cannot know that you have hands either. According to the skeptic, this point generalizes: whoever the epistemic subject *S* is, whatever the belief *p* is, and whatever the skeptical alternative is, *S* cannot know that *p* because, to know that *p*, *S* should be able to exclude all skeptical alternatives to *p*, but *S* cannot exclude any of its skeptical alternatives.

In the face of the skeptic's arguments, we can defend talk about knowledge by drawing attention to the role and importance of context. Normally, we know that we have two hands, the earth is round, and 2+2 equals 4. However, when skeptical alternatives are made salient to us, we no longer know those things. Before you started to read this Section 2, you were not even aware of the possibility of your being a BIV, and so you knew that you have hands *in that context*. However, by introducing in this section the skeptical alternative that you might be a BIV, the context has changed. Now, in this new context, the skeptical alternative that you might be a BIV has become salient. Now you realize that you cannot exclude the possibility of the skeptical alternative, and as a result, you no longer know that you have hands *in this context*.

Kindly note that changing the context, making mere logical possibilities salient to cast doubt, is exactly what the “merchants of doubt,” to whom we alluded at the opening of this chapter, typically do to subvert the call for political or legislative action. For instance, on behalf of tobacco companies, they argue that, despite appearances, it’s conceivable and thereby possible that there is no relation between smoking and lung cancer. Or they fabricate so-called scientific controversies to make the logical possibility salient that, since the industrial revolution, there has been no causal relation between human intervention and climate change (Oreskes and Conway 2010).

We can also defend ourselves against the skeptic’s argument by appealing to context in a slightly different way. We justify our claims all the time by adding reasons for these claims. To adopt a famous example from Austin (1946), when someone claims that a bird is a goldfinch, she might refer to its red head as typical of such a goldfinch. In response, someone else may wonder whether the bird is not a woodpecker instead of a goldfinch, pointing out that woodpeckers too have red heads. This is what we call an appeal to a relevant alternative. In the context of a bird in your garden, it makes sense to discuss whether it is a woodpecker or a goldfinch, but it would be really strange if someone wondered whether it was not a mechanical or stuffed bird, because such alternatives are not relevant in the context. So, to make a knowledge claim, an epistemic subject should only be able to exclude all relevant alternatives, and which alternatives are relevant is determined by the context. However, there are hardly any contexts in which skeptical alternatives are relevant. As a result, skeptical alternatives are hardly ever legitimately challenge knowledge claims.

#### 4. AGRIPPA’S TRILEMMA

**W**e might think that we can defend ourselves against wholesale skepticism and the conclusion that we really know nothing by arguing that the alternatives articulated by the skeptics are just very unlikely, that we do not have to take them seriously. There are several explanations for the things we experience, and some are better than others. It is on the basis of the quality of the explanations that we can discriminate between what to believe or not. Pyrrhonian skeptics, though, like Pyrrho, Sextus Empiricus, and Montaigne, have argued that there is no independent, unbiased, rational way to identify the best among competing criteria. Competing explanations simply have their own pros and cons. It is impossible to weigh these arguments properly, let alone settle for an independent, unbiased, rational conclusion. So, basically, Pyrrhonian skeptics are skeptical about the possibility of justification, and one of their main arguments to that effect is known as Agrippa’s trilemma.

Beliefs can be justified by further beliefs. In Austin's example, the birdwatcher's belief that the bird in the garden is a goldfinch is justified by her belief that goldfinches have red heads and that the bird at issue has the relevant characteristic. But what justifies these further beliefs? Still further beliefs in a justification chain? This can't go on ad infinitum, can it? According to Agrippa's trilemma, justification of beliefs by other beliefs is impossible altogether because there are only three options, and none of them is worth wanting: either

1. The justification chain terminates in basic beliefs that do not need further justification because they are, in a sense, self-justifying, like axioms in mathematics, or
2. The justification chain is circular, or
3. The justification chain does go on ad infinitum.

To avoid Pyrrhonian skepticism about justification, one might accept one of the horns of the trilemma. The first option is to accept the axiomatic horn. Throughout the history of ideas, by far most philosophers have adopted this approach, which is called *foundationalism*: like a pyramid, human knowledge in general and science in particular rest on a solid foundation of absolutely certain, infallible, basic beliefs. Consider perceptual knowledge. Of course, we may err; our perceptual judgments may be illusory or hallucinatory. However, we cannot be mistaken about having the pure sensory experiences that we do have. Our interpretations may be wrong, but we are infallible in having experienced what Descartes called sensations, Locke ideas, Hume impressions, Mill phenomena, and Russell sense-data. Beliefs that are ultimately justified by such basic beliefs are safe and sound.

There are many objections to classical foundationalism, but the main source of concern is the gap between, on the one hand, pure, uninterpreted sensory experiences and, on the other hand, minimally interpreted perceptual beliefs. Pure, uninterpreted sensory experiences may be infallible, but they are inferentially sterile; because they are experiences, and not yet propositional beliefs, nothing can be inferred from them, and hence they cannot provide foundational justification. Looking at a banana, I have a visual experience, which doesn't justify anything as such. Only when sensory experiences are minimally interpreted, i.e., if they are turned into propositional beliefs, can they provide justification. So, my visual experience may trigger me to form the belief that there is a curved yellow object in front of me, which in turn may provide justification for a further belief that the object in front of me is a banana. But the trouble with minimally interpreted perceptual beliefs is that they are no longer infallible – the object may in actual fact be pale green and straight – and hence, they cannot be the basic beliefs that terminate the justification regress either.



A second option is to accept the circular horn of Agrippa's trilemma. This results in the position we call "coherentism." According to coherentists, justification is holistic and not transferred from one belief to another. Beliefs hang together in more or less coherent "webs of beliefs." Whether or not a particular belief is justified is determined by whether and to what extent it fits or coheres with the other beliefs in the web of beliefs of the epistemic subject. The minimal conception of coherence is logical consistency, but most coherentists believe that explanatory connections contribute significantly to the overall coherence of beliefs and webs of beliefs. Intuitively, four completely unrelated beliefs are less coherent than, for instance, a doctor's belief that a single diagnosis explains three disparate symptoms; the diagnosis pulls the symptoms together.

There are many objections to coherentism as well, but the main source of concern is the isolation problem: coherent webs of beliefs can become completely detached from reality. Think, for instance, of conspiracy theories; those often make up internally coherent wholes, in which beliefs hook up neatly. However, from the "outside," people see that such theories are way off-base.

A third and final possibility is to accept the regressive horn of Agrippa's trilemma: the justification regress is not vicious, but virtuous. Beliefs can be justified by an infinite chain of justification. In response to the objection that the human mind is finite and limited, Klein (1999) argues that the infinitist is not claiming that the epistemic subject actually needs to have an infinite chain of reasons in mind. What is required is that it must be possible in principle, but not in practice, to provide an infinite chain of reasons. Moreover, Atkinson and Peijnenburg (2017) prove that the longer a justification chain becomes, the more the returns of further justification diminish, so that ultimately, no further justification is called for.

## 5. RELIABILISM

Up until now, we have considered the justification of beliefs by other beliefs. An underlying assumption has been that, at least in principle, the epistemic subject has access to the justification of a belief: the foundationalist can trace back or reduce her belief to a foundational layer of basic beliefs ultimately supporting it, the coherentist is or can become aware that her belief hangs together with other beliefs constituting a coherent web of beliefs, and so on. What these approaches to justification share is *internalism*: factors internal to the epistemic subject determine whether her belief is justified. However, triggered by the Gettier problem (see Section 1), philosophers have also developed externalist approaches to justification. According to these approaches, only factors external to the epistemic subject determine whether her belief is justified. According to

externalism, then, the epistemic subject needs no access to the justification of a belief.

Reliabilism is such an externalist theory of justification. To grasp the basic notion of reliability, we can think of the debate about the media and “fake news.” Journals or networks that often publish news that turns out to be false are unreliable. Readers, listeners, or viewers should not rely on them, but rather base their beliefs on information obtained from reliable outlets: journals and networks that almost never publish false information. So, in this context, being reliable is a function of being truth-conducive; the degree of reliability of the media depends on the ratio of the true to false beliefs they tend to produce. Goldman (1979) argues that what is true for networks and journals is also true of ourselves: some of the ways in which we acquire beliefs are clearly less reliable than others. Guesswork, hasty generalizations, and wishful thinking are clearly unreliable because, more often than not, they produce false beliefs. In contrast, belief-forming processes like standard perceptual processes and good reasoning are reliable: they tend to produce more true than false beliefs. Hence, reliabilism proposes that a belief is justified if and only if it has been formed by a reliable process of belief-formation, and where the reliability of the process is defined in terms of its tendency to form beliefs that are true rather than false.

Unfortunately, reliabilism can be challenged by the Brain-In-a-Vat version of the evil demon scenario discussed above. The reason for that is that we need to be sure about the existence of the external world and our (perceptual) relation to it before we can say anything about the reliability of our perceptions. Suppose that in some scenario, your counterpart, i.e., your local representative in some possible world, has exactly the same perceptual experiences as you have, and forms exactly the same perceptual beliefs as you do. However, there is one all-important difference: in contrast to you, your counterpart is systematically misled by an evil demon and is in actual fact a Brain In a Vat. As a result, almost all your counterpart’s perceptual beliefs are false, and we must qualify the perceptual processes by which they are formed as highly unreliable. However, the input is the same: you and your counterpart have the same perceptual experiences. And the output is the same as well: you and your counterpart form the same perceptual beliefs. So, if your perceptual beliefs are justified, your counterpart’s perceptual beliefs should be justified as well.

There are several other problems with reliabilism that we will not discuss here. However, one thing to note is that a major advantage of reliabilism is that, unlike internalism, it allows for what Sosa calls animal knowledge. A significant part of human knowledge consists of first-order, reliably produced true or apt belief. Your first-order true belief that you are now reading this chapter is reliably formed based on perception (and perhaps also a reliable bit of introspection). However, according to Sosa’s own hybrid account, human knowledge is not restricted to

animal knowledge: you can upgrade your apt belief to what Sosa calls reflective knowledge by reflecting on how your apt belief fits coherently with other beliefs you have, like, for instance, your belief that you have bought this book in that bookshop, your belief that you could hardly wait to start reading this chapter, and so on.

## 6. THE CONTEXT OF INQUIRY

**W**e started this chapter with a handful of pressing and complex real-world problems, all related to the epistemic quest of finding out what is and isn't true. Subsequently, we discussed big questions in epistemology – the analysis of knowledge, the possibility of knowledge, and the nature of justification – and to that end, we made abundant use of imaginary examples and counterexamples. At face value, we are now facing the applicability gap, i.e., the gap between, on the one hand, deep but narrow academic puzzles, and, on the other hand, complex real-world problems. Now, in what sense and in what way can philosophy in general, and epistemology in particular, help to bridge that applicability gap? What modest role can philosophy and epistemology play in addressing and solving the real-world problems of our time?

The first key to reaping the harvest of more than 25 centuries of philosophy, for the purpose of solving pressing real-world problems, is “context.” Philosophical questions often are decontextualized questions, and that is why, ever since Plato, stipulated, imaginary cases and scenarios are the ideal tools for philosophical investigations. The downside is that there never seem to be definite answers to decontextualized questions. Philosophers prefer to put more and more possible solutions on the table, and to go on and on with considering and weighing their pros and cons. In a contextual approach, however, one tries to enrich the definition of a problem by bringing in more details and context in such a way that, at the end of day, a solution can be arrived at, i.e., the best solution for the problem at hand in its context.

In addressing real-world problems, philosophical, conceptual questions often and easily pop up. For instance, to become clear about what one knows at some point in time, one needs a conception of knowledge. Or to take another example, to assess to what extent a theory is justified, one needs a conception of justification. As we can learn from Sections 1, 3, and 4, without context, the analysis of concepts like “knowledge” or “justification” or still, for that matter, “coherence,” is an endless game. There simply are no one-size-fits-all solutions to philosophical, conceptual problems because in different contexts, different conceptions are appropriate. Only when we are clear about the context of the real-world problem that we want to solve, our goals, and so on, can we start defining a concept in a fruitful way. To that end, the exploration of possible conceptions

during 25 centuries of philosophy has not been in vain. On the contrary, it allows the real-world problem-solvers to significantly speed up defining the concept at hand for the purpose at hand.

From our discussion of skepticism about knowledge in Section 2 and of skepticism about justification in Section 3, we can learn a second important lesson in how epistemology could and should make itself more useful in addressing and solving real-world problems. Epistemology should reorientate itself and focus more on inquiry, how to obtain knowledge, and how to find out what is and what isn't true, than on analyzing end-products. In the real world, epistemic subjects in general, and problem-solvers in particular, are always in the midst of inquiry: some things they can safely assume and act on, whereas other things they need to find out or investigate further. In the midst of inquiry, there is little use for analyses of idealized, unattainable goals. By contrast, a profound understanding of the structure of inquiry, i.e., the structure of problem-solving itself, would be most useful. For instance, it would suggest to problem-solvers what the best next step is in solving the complex, real-world problem they are addressing.

Let me illustrate this point by adopting an analogy from Schurz and developing it a little bit. Suppose that someone in a hurry asks an epistemologist for the right way to the railway station. A traditional epistemologist will respond: "Among all possible ways that start where you are right now and end at the railway station, find the shortest one and choose it." With that answer, the epistemologist does not risk anything. It is a true answer, but also completely useless. It bounces back the question that the person asking it is interested in, acting as if it is an answer. Epistemologists can do so much better, but not by providing substantive answers – after all, it is their profession to think and reflect on what knowledge is, not provide answers to all possible questions people might have. What they can do is provide what we can call a heuristic answer, i.e., an answer that sheds light on how to find the answer to the substantive question. The epistemologist can shed light on the quickest way to find that answer, or point out that one should first get one's goals clear, and so on.

These two lessons fit together nicely. As we discussed toward the end of Section 2, "contexts" or problem-solving situations not only allow for fruitful conceptions but also determine which alternatives are relevant. In their turn, the relevant alternatives determine the further steps to be taken and the further questions to be asked in the process of solving the real-world problem. As a matter of principle, all voices in the marketplace of ideas deserve to be heard, but the epistemologist can learn and teach how to single out the most relevant ones, step by step.<sup>1</sup>

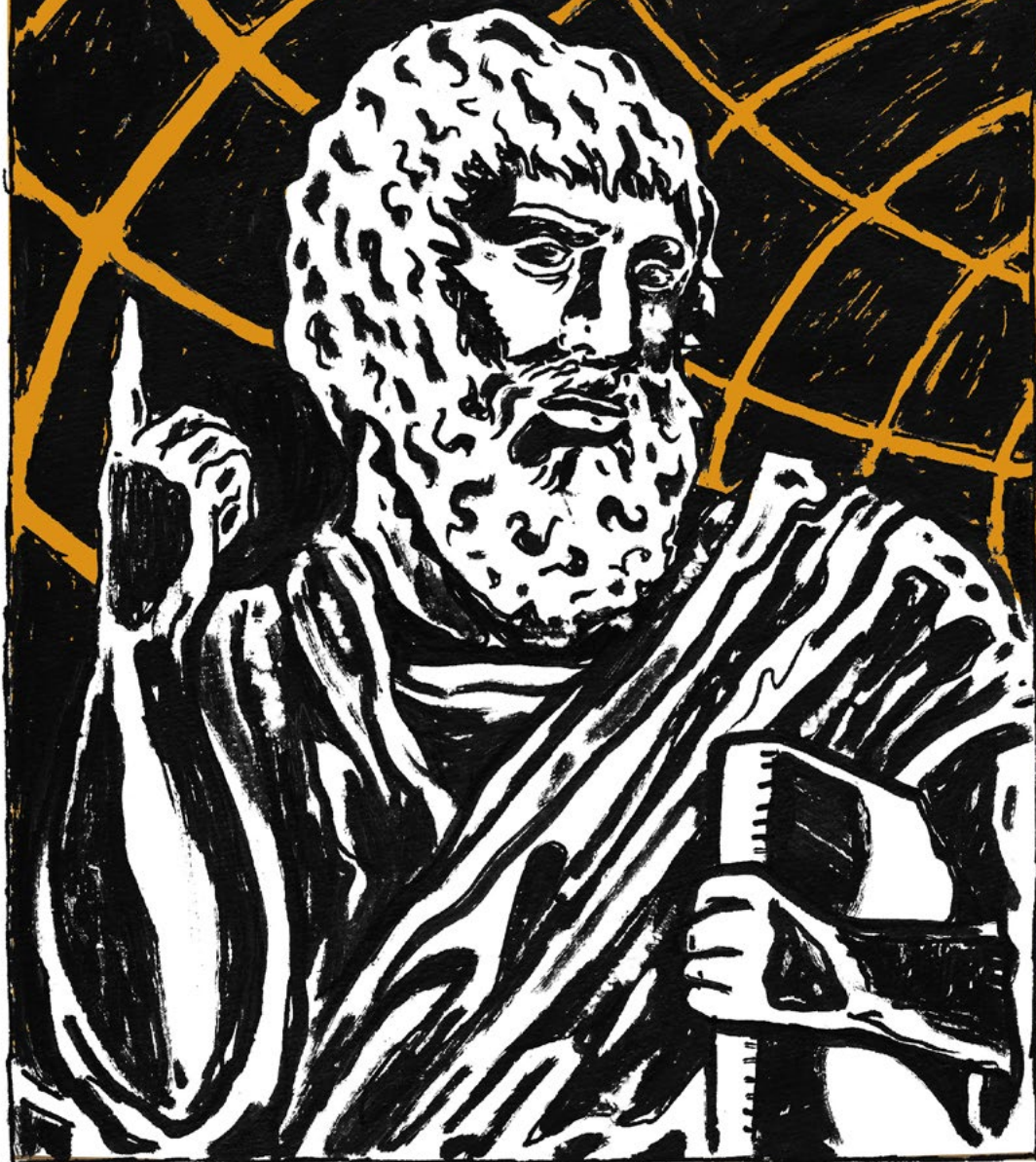
---

1. I am very grateful to Maureen Sie, Bart Engelen, Lodi Nauta and Martin van Hees for their insightful comments on earlier drafts of this chapter, and for their editorial patience and dedication.

## BIBLIOGRAPHY

- Adriaenssen, H. T. (2013). *Knowledge and the veil of representations* [PhD Dissertation]. University of Groningen.
- Atkinson, D., & Peijnenburg, J. (2017). *Fading foundations: Probability and the regress problem*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-319-58295-5>
- Austin, J. (1946). Other minds. *Proceedings of the Aristotelean Society, Suppl. Vol. 20*, 148–187.
- Dretske, F. (1971). Conclusive reasons. *Australasian Journal of Philosophy*, 49(1), 1–22. <https://doi.org/10.1080/00048407112341001>
- Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23(6), 121–123. <https://doi.org/10.1093/analys/23.6.121>
- Goldman, A. (1967). A causal theory of knowing. *The Journal of Philosophy*, 64(12), 357. <https://doi.org/10.2307/2024268>
- Goldman, A. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, 73(20), 771. <https://doi.org/10.2307/2025679>
- Goldman, A. (1979). What is justified belief? In G. S. Papas (Ed.), *Justification and knowledge* (pp. 1–23). Springer Netherlands. [https://doi.org/10.1007/978-94-009-9493-5\\_1](https://doi.org/10.1007/978-94-009-9493-5_1)
- Harding, S.G. (1991). Whose science? Whose knowledge? In *Whose science? Whose knowledge? Thinking from women's lives*. Cornell University Press. <https://doi.org/10.7591/9781501712951>
- Klein, P.D. (1999). Human knowledge and the infinite regress of reasons. *Nous*, 33(s13), 297–325. <https://doi.org/10.1111/0029-4624.33.s13.14>
- Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.
- Oreskes, N., & Conway, E. (1981). *Merchants of doubt*. Bloomsbury.
- Russell, B. (1948). *Human knowledge: Its scope and limits*. Routledge. <https://doi.org/10.4324/9780203875353>
- Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge*. Oxford University Press.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

PLATO



## PLATO *THEAETETUS*

---

Plato's *Theaetetus* heralds the start of almost 25 centuries of heated philosophical debates about the nature of knowledge. Oversimplifying a bit, the dialogue successively considers the analyses of knowledge as perceptual, true, and justified true belief. It will likely not satisfy readers because all three conceptions are found wanting. Nevertheless, it's very rich and exhibits all the great features of Platonic dialogues, especially Socrates' insightful cleverness, making him raise deep questions about and powerful objections to the conceptions and views being considered. For instance, against Protagoras' famous relativist statement that "man is the measure of all things," Socrates develops his turning-around or reversal argument: relativism is only intelligible if it excludes non-relativism, but the exclusion of non-relativism boils down to the denial of relativism.

Although the substantive points are very interesting and important, Platonic dialogues in general and the *Theaetetus* in particular also pave the way methodologically. They pioneer the use of imaginary cases or "thought experiments" in testing philosophical conceptions and theories. Instead of restricting one's attention to real-world examples and counterexamples in all their complexity, philosophical methodology allows and even recommends using one's imagination, especially in inquiring what the nature or essence of a philosophically relevant phenomenon like "knowledge" really is. For instance, the second part of the *Theaetetus* is settled with a relatively straightforward counterexample to the conception of knowledge as true belief: an imaginary jury passing an accidentally true judgment about the guilt of someone accused of theft, based only on hearsay. It would be counter-intuitive to say that this jury knows that the accused is guilty of theft.

---

Chris Buskes

**The Bounds  
of Reason**



# THE BOUNDS OF REASON

## *An Introduction to Philosophy of Science*<sup>1</sup>

Chris Buskes  
[c.buskes@ftr.ru.nl](mailto:c.buskes@ftr.ru.nl)

### 1. INTRODUCTION: WHAT IS SCIENCE?

**B**etween the 1960s and 1990s there was a heated debate in the United States about 'Equal Time', a claim put forward by fundamentalist Christians who demanded that public schools should teach creationism alongside evolution in science class. The fundamentalists' plea was that the same amount of time should be spent on *both* tenets – hence 'equal time' – so that students would be able to choose for themselves what to believe. The debate culminated in several renowned lawsuits, up to the Supreme Court. Creationism is the doctrine that the Biblical account about the origin of the universe, the Earth, and man is literally true – and that Darwin's theory of evolution and other sciences which are at odds with creationism like geology, paleontology, cosmology, archaeology and so on are all plain wrong. After all, the Book of Genesis tells us that everything was created by God somewhere within the last 10,000 years.

Opponents of the idea of Equal Time argued that an introduction to various competing ideas and worldviews would be fine in itself but that, unlike the theory of evolution, creationism is not a *scientific* theory and therefore has no place in science class. Not surprisingly, creationists did not agree and insisted that their

---

1. I would like to thank Lodi Nauta, Martin van Hees and Bart Engelen for their helpful comments on an earlier draft of this chapter.

worldview, they now called it 'creation science', is just as scientific as Darwin's theory of evolution. However, in the early 1980s several court rulings decided otherwise. In 1981 a judge in Arkansas, William Overton, ruled that creationism is not science because, among other things, it is not testable. Creationism was therefore qualified as 'pseudoscience'. That is, creationism *pretends* to be science but since it doesn't conform to certain scientific criteria and methodological standards, it isn't actually science at all. Although the ruling by judge Overton did not apply to public schools outside Arkansas, it nevertheless had a considerable impact on subsequent court rulings in the US with regard to teaching creationism in science class (Bowler, 2009). It seemed that science had prevailed, at least for the time being.

The debate about Equal Time raises several interesting questions about the nature of science and scientific knowledge. When is it appropriate to call an idea or a theory 'scientific' and when is it not? What exactly demarcates science from non-science and pseudoscience? Is science always a purely logical and rational enterprise or can its content and development be influenced by extra-scientific and irrational forces? What *is* science for that matter? These are the kinds of questions that philosophy of science, as a sub-discipline in philosophy, is raising and trying to answer.

This chapter will mainly focus on the nature of the natural sciences but much of the analysis also pertains to the social sciences. We begin with a section about the so-called 'demarcation problem', the question what separates science from pseudoscience. In subsequent sections we will address the socio-historical turn in the philosophy of science, the rise of relativistic and postmodern views on science, and the reaction this in turn provoked. Via the so-called 'science wars', the clash between diametrically opposed perspectives on science, we end this chapter with the question whether such completely different views can be reconciled or not. The recurring theme is the question: what is the nature of science?

## 2. THE DEMARCATION PROBLEM

**A**t first glance the issue seems quite simple. What separates science from non-science and pseudoscience is that science is empirically testable. Science is concerned with what can be observed and measured, and for which empirical evidence can be gathered. Scientific theories generally make predictions which can be tested in reproducible experiments. All this makes science a self-corrective enterprise since it is always open to critique and revision, for instance if new facts are discovered.

In the first half of the 20<sup>th</sup> century representatives of a dominant movement in philosophy at the time called 'logical positivism' argued that verifiability is what separates science from pseudoscience. More precise, they held that there are two kinds of meaningful statements: analytic statements (as in mathematics and logic), and statements that can be verified by experience, i.e., of which we can tell whether they are true or false. A (non-analytical) statement which cannot be verified is considered meaningless, like the proposition "The supreme being is benign" or "The soul is immortal". The logical positivists were strongly inspired by the empiricist tradition of David Hume, Auguste Comte, Ernst Mach and others which holds that all knowledge is empirical in nature and thus grounded on sense experience. So as true empiricists they believed that verifiability is the solution to the demarcation problem: theories which can in principle be verified are scientific, whereas theories which cannot be verified are pseudoscientific or just nonsensical. Thus the criterion of verifiability was believed to bring about one of the main goals of logical positivism: it shows that metaphysics and all other kinds of 'obscure' thinking are thoroughly unscientific (Ayer, 1936).

Yet on closer inspection the demarcation problem turned out to be more persistent. For instance, it soon became clear that the verification principle was too stringent because it not only labels pseudoscience as meaningless but also significant chunks of genuine science like natural laws and other universal generalizations. After all, a statement like "All ravens are black" cannot be conclusively verified because it is impossible to check all ravens in the present, past and future.

In philosophy this is known as the problem of induction. Even if all ravens *which we have observed until now* were black, we can never be certain that *all* ravens are black. We cannot rule out that maybe tomorrow, or in ten years, a non-black raven pops up (or that the sun will rise in the west, to take another example). Some logical positivists therefore suggested that the criterion of verifiability should be adjusted and transformed into a criterion of confirmability. Thus although the statement "All ravens are black" (or "The sun always rises in the east") can never be conclusively verified, it can be confirmed. After all, every observed black raven is a confirmation of the hypothesis that all ravens are black. Yet whereas the criterion of verifiability was too stringent, the criterion of confirmability turned out to be too weak because it opens a crack through which pseudoscience can sneak back in again. Pseudoscientific statements can often be easily confirmed because pseudosciences generally do not make precise and bold predictions – and the more vague a prediction is, the easier it is confirmed. Think for instance of a horoscope that says: "Your patience will be put to the test in the coming week." While this prediction can be confirmed, astrology is not a science.

The philosopher of science Karl Popper, a critic of logical positivism, therefore introduced the criterion of falsifiability as the solution to the demarcation

problem. The flaw of pseudoscience, Popper (1959) argued, is that it cannot be falsified because it covers itself against critical scrutiny. Thus a pseudoscience like astrology can easily be *confirmed* because its claims and predictions are so vague, but for the same reason it cannot be *refuted* because astrologists do not stick their necks out. Thus only theories which are falsifiable, which can clash with possible observational data, are genuinely scientific.

Furthermore, given the problem of induction, no amount of observations can ever conclusively *prove* a hypothesis, whereas one single counterexample can *disprove* it. Put differently, we can never know whether a theory is true, but we can know whether it is false. Thus by tracing the faults and shortcomings in our theories we can gradually improve our knowledge. Popper's suggested solution to the demarcation problem may look plausible, but on closer inspection it turns out that it has its own weaknesses. For instance, contrary to what Popper argues, alleged pseudosciences like astrology are indeed falsifiable, for instance the astrologist's claim that zodiac signs correlate with certain character traits. These claims can be and have been proven false (Carlson, 1985). So astrology is not only falsifiable but indeed actually *falsified*, and according to Popper's own criterion that means that astrology is science – *refuted* science, but science nonetheless.

Philosophers of science like Imre Lakatos, a former disciple of Popper, and Larry Laudan argue that all attempts to solve the demarcation problem with one simple criterion are doomed to fail. Moreover, we cannot determine at any given time, as in a snapshot, whether a particular discipline, or 'research programme' as Lakatos (1970) calls it, is genuinely scientific or not. It is only in the long run that we can tentatively assess a particular discipline's merits and shortcomings. This means that the difference between science and pseudoscience is not as straightforward as we initially might have hoped. There is no clear boundary between the two domains because they may partially and temporarily overlap. Science has no eternal, unchanging essence. Some philosophers of science therefore concluded that the demarcation problem is unsolvable in principle (Laudan, 1996, chapter 11). Moreover, it seems that we cannot determine once and for all what characterizes scientific knowledge because our ideas about what science is may change over time.

### 3. SCIENCE, HISTORY AND RATIONALITY

**A**n observant reader might object that although our ideas about science might change over time, one particular feature of science probably will never change, namely its rationality. Thus irrespective of its content, science is and always will be a rational enterprise because its method of

inquiry is determined by logic and reason. By pursuing objective truth, science successfully manages to transcend social and historical contingencies.

Before we delve into this intricate matter, it is useful to make a distinction between what is called the 'context of discovery' and the 'context of justification'. That is, scientific inquiry consists of two different aspects, the first being the *generation* of new ideas and hypotheses, and the second the subsequent *evaluation* of these new ideas and hypotheses. Most contemporary philosophers of science argue that only the latter process is rational and guided by logic and reason. We *do* have logical rules and rational procedures for determining whether a new idea is justified, but we have no such rules and procedures for how to conceive a new idea, because this process is largely intuitive and therefore non-rational.

But there is much more to say about the alleged rationality of science. In the 1960s and 1970s several radical philosophers of science emerged that rebelled against the traditionalist view of Popper and the logical positivists. These philosophers included Thomas Kuhn and Paul Feyerabend, among others. Because these thinkers emphasized the historical context of scientific practice, they are sometimes called the 'historicists' (Nickles, 2017). These philosophers deviated from the traditional approach in that they question the supposed universal rationality and objectivity of science. Instead they defend a more relativistic view in which truth seems contextual, or relative to a conceptual scheme (I deliberately use the word 'seems' here because there are multiple interpretations of Kuhn's and Feyerabend's philosophies, both moderate and radical). Simply put, what people believe to be true and real always depends on a particular historical, social or cultural context. And since we do not have an objective standard with which to compare different worldviews, it is impossible to say which one is best.

Thus Kuhn (1962) speaks of different 'paradigms' in the history of science, each of which has its own rationale. A paradigm is a scientific worldview which dominates a particular discipline for a certain amount of time (during what Kuhn calls 'normal science') until it is discarded for a new paradigm in a scientific revolution. Examples of such paradigm changes are Copernicus' heliocentric astronomy which superseded the geocentric worldview of Aristotle and Ptolemy, or Einstein's theory of relativity which displaced classical Newtonian physics. Kuhn's controversial claim is that paradigms are 'incommensurable', i.e., they cannot be objectively compared because there is no neutral standard available. According to Kuhn, observation is 'theory-dependent' or 'theory-laden', which means that observation cannot serve as a neutral arbiter among competing theories because what we see depends (to some extent) on what beliefs we hold and which worldview we embrace. So we cannot appeal to neutral, objective facts because ultimately facts are always relative to a particular conceptual scheme. Kuhn (1962, p. 111) therefore concluded that "when paradigms change,

the world itself changes with them". According to Kuhn, a change of paradigms is like a 'gestalt switch' which causes scientists to see the world in a completely different light.

Moreover, Kuhn argues that the revolutionary transition from one paradigm to the next is not a well-thought-out rational process because there is no objective way of measuring a paradigm's success. Thus a paradigm shift is more like a religious conversion and consequently the development of science can no longer be viewed as a cumulative and rational process which brings us ever closer to the truth. Although in a Postscript to the second edition of his book, Kuhn (1970, p. 206) denies the accusation of being a relativist, his views are still generally interpreted as tending towards relativism. It should be noted that this discussion has considerable societal relevance, especially in our current 'post-truth' era in which people talk of 'alternative facts' and where relativism has made its way into popular opinion.

An even more radical tendency toward relativism and irrationalism can be found in the works of Paul Feyerabend, the maverick among 20<sup>th</sup> century philosophers of science. Feyerabend (1975) calls himself a 'methodological anarchist' because he believes that there is no unchanging, universal scientific method which supposedly governs all instances of exemplary science. Instead, he urges us to try out lots of different methodologies under the motto 'anything goes'. Moreover, Feyerabend believes that in order to broaden our horizon we should even actively go *against* established methodologies because by breaking the rules we might open up new vistas.

In addition, Feyerabend pleads for theoretical pluralism because he thinks that the complexity of the world can never be captured by a single approach or in a single perspective. Consequently, Feyerabend argues that (Western) science is only but one belief-system among many others. Alternative belief-systems and traditions like voodoo, Chinese medicine and Indian rain dance rituals, may provide us with knowledge as well and are therefore equally respectable. Feyerabend, who defended astrology, thus flatly denies that there is a clear-cut distinction between science and pseudoscience.

#### 4. SCIENCE IN THE MAKING

**T**he works of Kuhn, Feyerabend and other anti-traditionalist thinkers paved the way for a new approach in philosophy of science which took the study of the socio-historical dimension of science to a next level. The rise of the sociology of science and postmodern theories of knowledge in particular played an important role in this development. Sociologists of science have underlined the socially-conditioned nature of knowledge, while postmodernists have argued

that the ideas of objective knowledge and absolute truth are illusory. The notion of absolute truth is rejected in favor of the idea that there are only different points of view or 'narratives', each of which has no more claim for truth than any other. Knowledge claims always depend on a particular frame of reference and can therefore only be validated *within* that context of origin. The combination of these two doctrines – the sociology of science and postmodernist theories of knowledge – resulted in a rather bleak and radical form of relativism which holds that science is a social construct, and that its content and development are determined by interests, power and ideology.

Perhaps unwittingly, Kuhn and others triggered the so-called 'sociological turn' in the philosophy of science. The sociological turn refers to the idea that science must be studied and understood from a sociological point of view. That is, we should not look so much at 'ready made science' but rather at 'science in the making', as the sociologist Bruno Latour (1987) calls it. Ready made science refers to the end products of scientific labor such as published articles and books. But according to Latour and others, this is only the proverbial tip of the iceberg. Much more interesting is what hides below the surface, namely science in the making, i.e., what scientists actually *do* every day. After all, science is a social and collective enterprise.

So sociologists of science argue that we must shift our attention from ready made science to actual scientific practice. When we study scientists in their daily interaction, then, we soon find that facts are not discovered but 'constructed' during a complex process of negotiation, persuasion and trying to reach consensus. When consensus *is* achieved, the outside world is then told that "scientist X has discovered fact Y", thereby keeping silent about the complex social processes that preceded the 'discovery'. In other words, facts are not lying around out there in the world waiting to be discovered, but facts are rather made by us. They are *artefacts*. So it is not the facts that lead to consensus, but the consensus that leads to facts. Furthermore, in the negotiation process some individuals, classes or communities (e.g., white, western males) have more influence than others, which means that ultimately the content and development of science is sociologically determined and governed by skewed balances of power.

Sociologists of science like Barry Barnes and David Bloor, who founded the so-called 'Strong Programme' in the sociology of knowledge, believe that their inquiry is more fundamental than the issues raised by traditional philosophers of science of the pre-Kuhnian era. Instead of asking what demarcates science from pseudoscience, or what makes some beliefs rational and others irrational, we should rather ask *why* people *believe* that something is science rather than pseudoscience, and *why* people *believe* that something is rational rather than irrational. According to Barnes and Bloor, these questions are more profound

because they point to the underlying and ultimate social *causes* of our beliefs and states of knowledge.

Thus someone may be inclined to call something 'pseudoscience' or 'irrational' because he or she has an interest in such stereotypical thinking, or perhaps it is someone's social background or education that makes him or her tend to pigeonholing. Suppose a traditional philosopher of science would say that it is rational to believe in, say, astronomy and molecular biology and irrational to believe in homeopathy and parapsychology because the former are sciences and the latter pseudosciences. A sociologist of science then could riposte that all such labeling might well be arbitrary because one's preferences and prejudices are themselves socially determined. So, in the end, our task is not to judge but to unravel and understand the social causes of our beliefs (Bloor, 1976).

## 5. THE SCIENCE WARS

In the previous sections we saw that Kuhn, Feyerabend, and the sociologists of science all have emphasized the social and historical dimension of scientific knowledge. That is, what is held true and real is always relative to a particular conceptual scheme, a paradigm, or a worldview, et cetera. Extreme versions of this idea led to the postmodern and social constructivist view that there is no absolute truth, true knowledge or objective facts because truth, knowledge and facts are all contingent social constructs. This relativistic view thus assumes that the formation of our beliefs is not constrained by the properties of an objective world. However, this claim is not without controversy.

In the 1990s, the popularity of social constructivism and postmodernism caused a reaction from scientists and philosophers who opposed this kind of radical relativism. It led to a fierce argument over the nature of truth and the alleged objectivity of knowledge and science. This clash of ideas became known under the name the 'science wars'. On the one hand there were the scientific realists and objectivists who maintained that objective knowledge and truth are real. And on the other hand there were the relativists and postmodernists who denied the existence of objective knowledge and truth.

In all fairness we can say that both camps have their weaknesses. The realists face the problem of human fallibility. We have often believed that something is true, but again and again it turned out we were mistaken. So some modesty is appropriate because eventually every theory will turn out to be false or at least incomplete. Given our fallibility then, we should accept that our knowledge is always tentative and conjectural. Or as Popper (1959, p. 278) once said: "We do not know, we can only guess." Yet despite our fallibility Popper believes that our knowledge is getting closer to the truth, which is illustrated by the impressive



success of modern science and all of its technological applications. We will return to this issue below. The radical relativists also face a problem because their claims are self-refuting since the proposition that states that “Everything is relative” entails that the proposition itself cannot be true either. The claims of the sociologist of science may face a similar problem. After all, if facts are social constructs, what does this mean for the ‘facts’ presented by sociologists of science themselves? Are those ‘facts’ in any sense true or trustworthy? It seems that the sociologists of science are caught up in some sort of paradox because their claims undermine their own thesis. Or as the philosopher Roger Trigg (2001, p. 37) puts it: “If the strong programme of the sociology of knowledge is true, it is thereby shown to be false.”

In 1996, the American physicist Alan Sokal (1996) submitted a paper to *Social Text*, a fashionable academic journal of postmodern cultural studies. The manuscript had the seemingly profound title ‘Transgressing the boundaries: Towards a transformative hermeneutics of quantum gravity’, consisted of a complex mix of physics and philosophy, and argued that physical reality must be viewed as a social and linguistic construct. Yet immediately after the article was accepted and published, Sokal revealed that he had in fact submitted a bogus paper full of gibberish. His real intention was to show that the journal’s editors and readers didn’t have the faintest clue of what science is actually about. Sokal, situated in the camp of the realists, had pulled a fast one on the postmodern relativists. The ‘Sokal hoax’, as it was soon dubbed, was one of the low points (although others probably would say highlights) of the science wars. A few years later Sokal wrote a book with his fellow physicist Jean Bricmont, called *Fashionable Nonsense*, in which they supposedly wiped the floor with all postmodern and relativistic trends in university departments (Sokal & Bricmont, 1998). But this publication caused much less uproar. Eventually the science wars subsided and tempers have calmed down, but we are still witnessing radically different perspectives on science which seem difficult to reconcile. So this issue is by no means resolved.

## 6. SCIENTIFIC REALISM

Scientific realism is characterized as “[.] a positive epistemic attitude toward the content of our best theories and models, recommending belief in both observable and unobservable aspects of the world described by the sciences.” (Chakravartty, 2017). Scientific realists argue that their opponents, the postmodern relativists and social constructivists, don’t really understand much about the true nature of science. Realists readily admit that science is a social enterprise and that its development can be influenced by external, socio-historical factors. But unlike other products of the human intellect such as art

and literature, science has one aspect that makes it completely different. That is, science can be *tested* against reality through observation and experiment.

Moreover, a realist argues that one cannot ignore the spectacular success of modern science. Think for instance of how modern medicine, and especially our knowledge about hygiene, pathogens and vaccination has saved and will save countless lives, thereby spectacularly increasing our life expectancy. Or think of the successes in molecular biology like the discovery of DNA and the sequencing and mapping of the human genome (and that of many other species). Or consider recent breakthroughs in physics and astronomy like the discovery of thousands of exoplanets, the detection of gravitational waves, and the first picture of a black hole. According to the realist, this eye-catching success is a clear indication that scientific theories and facts are not arbitrary social constructs. On the contrary, the best explanation for the success of science is the assumption that our current theories must be at least approximately true and that the entities to which they refer like genes, chromosomes, electrons, quarks and black holes do really exist. If these 'entities' were only figments of the socially-conditioned human mind, the success of science would be truly miraculous.

In addition, a scientific realist can argue that facts aren't merely social constructs because the objective world can prove even our most cherished beliefs to be wrong. For a long time we were convinced that the Earth occupies a special, privileged place in the center of the universe, and that a supreme being created all species separately a few thousand years ago. These and other fostered beliefs turned out to be false. So apparently the properties of the world put constraints on our theorizing. Even if we passionately want something to be true, the world may nevertheless decide otherwise (Psillos, 1999).

A reader might object that the claim that our current theories must be approximately true is at odds with the premise of human fallibility. As we mentioned earlier, it is difficult to maintain that our current theories are true because sooner or later every theory will be adjusted or displaced by a new, and hopefully better, theory. After all, that is the lesson the history of science teaches us (Laudan, 1981). So perhaps the scientific realist should be a bit more modest and reserved in his or her claims. A more cautious realist might therefore say that our current theories may indeed *not* be approximately true, but that the entities to which these theories refer do actually exist, even when they are unobservable, because we can detect and manipulate them with the help of all sorts of instruments. So although our theories about, say, electrons may turn out not to be approximately true, electrons *do* exist because we can manipulate them, for instance in a laboratory (Hacking, 1983). In any case, weak or strong, bold or modest, what these varieties of realism have in common is that they treat scientific theories as statements about an objective world, statements which are therefore true or false, and that can be tested through observation and experiment. The result is

that our knowledge about the world gradually increases, and according to the scientific realist this means that science is not an arbitrary social construct.

## 7. PERSPECTIVES ON SCIENCE

**N**ow let's return to the opposition against Darwin's theory of evolution. The debate offers an interesting case study to which our knowledge acquired in this chapter perhaps can be applied. From the 1990s and onwards, antievolutionists across the world (but mainly in the United States) have put forward a new weapon in their battle against Darwin, to wit: Intelligent Design (ID). ID no longer refers to the Book of Genesis, nor does it openly defend any particular faith. Instead proponents of ID argue that the theory of evolution is fundamentally flawed because it cannot explain the origin of 'irreducibly' complex design. Advocates of ID assert that all instances of complex design presuppose an intentional agent, an intelligent *designer*, just as a watch needs a *watchmaker* to come into existence. Furthermore, proponents of ID argue that, unlike creationism, ID is a genuine scientific theory and a worthy competitor to Darwin's theory of evolution. ID can and should therefore be taught in public schools during science class without raising any suspicion about promoting religion.

Now what should we think about this new strategy? Is ID indeed science, or rather pseudoscience? The philosopher Philip Kitcher (2007) has argued that it is neither. He qualifies ID as 'dead science' because ID simply tries to resurrect an age-old idea – the so-called teleological argument – which Darwin already refuted more than 150 years ago. Of course, advocates of ID aren't impressed with this qualification. In their view ID is indeed proper science. It is rather Darwinism which must be considered a pernicious philosophy. So the debate and controversy about evolution continues, at least in some peripheral academic circles. A 2017 Gallup Poll shows that, although reaching a new low, still 38% of Americans reject evolution outright and endorse the belief that God created human beings in the present form at one time within the last 10,000 years (Swift, 2017). So the battle over science class curriculum isn't over yet. What this case demonstrates is that debates over what counts as science (in the eyes of the public) and what should be taught in schools as science remains a contentious matter. It also demonstrates, again, the societal relevance of philosophy of science. We need to think hard about what science is and how it works if we want these debates to be decided on something else than pure 'power politics' or ideology.

Now let's take stock. In this chapter we have explored the nature of science and scientific knowledge. We began our inquiry with a discussion of the demarcation problem and the (unsatisfactory) solutions that Popper and the logical positivists

have offered to it. We then addressed the socio-historical turn in philosophy of science initiated by Kuhn and Feyerabend among others. Eventually we ended up with two diametrically opposed views on science, one from the relativists and constructivists and the other from the realists and objectivists. These two perspectives are the opposite extremes of a broad spectrum which also contains many intermediate positions. One such intermediate position perhaps could reconcile the two views to some extent. That is, one could acknowledge the fact that science is indeed in some sense a 'social construct' because all science is done by us, human beings. And since we humans are fallible, we should always be beware of hubris. Yet at the same time one can maintain that an objective world puts constraints on our theorizing and that current theories are closer to the truth than previous ones. So we might steer a middle course between self-defeating forms of relativism and overconfident forms of realism, thereby avoiding the drawbacks of the two extremes (Haack, 2003).

Another conclusion is that our ideas about the nature of science, and about its scope and limits, change over time. The boundaries of science are fluid. What this means is that there is no way to determine once and for all what science is. In fact, because a *general* theory of science turned out not to be feasible, the different branches of science (i.e., the natural sciences, the life sciences, the social sciences, the humanities) now all have their own *particular* philosophy of science. So instead of a single philosophy of science we now witness various *philosophies* of the various sciences (Allhoff, 2010). Because we cannot determine once and for all what the designation 'scientific' exactly means and when it is legitimate, we should rather assess it case by case. And since there is no external vantage point or first philosophy, we have no choice but to use the findings of science itself in our never-ending quest to gauge the bounds of reason (Quine, 1969).

## BIBLIOGRAPHY

- Allhoff, F. (Ed.). (2010). *Philosophies of the Sciences*. Wiley-Blackwell.  
<https://doi.org/10.1002/9781444315578>
- Ayer, A. (1936). *Language, truth and logic*. Penguin Books.
- Bloor, D. (1991). *Knowledge and social imagery* (2nd rev ed.). The University of Chicago Press. (Original work published 1976)
- Bowler, P. (2009). *Monkey trials and gorilla sermons: Evolution and Christianity from Darwin to intelligent design*. Harvard University Press.
- Carlson, S. (1985). A double-blind test of astrology. *Nature*, 318(6045), 419–425.  
<https://doi.org/10.1038/318419a0>
- Chakravartty, A. (2017). Scientific Realism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017, pp. 225–236). Stanford University Press.  
<https://doi.org/10.4324/9780203712498-19>
- Feyerabend, P. (1975). *Against method*. Verso.
- Haack, S. (2003). *Defending science within reason: Between scientism and cynicism*. Prometheus.
- Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.
- Kitcher, P. (2007). *Living with darwin: Evolution, design, and the future of faith*. Oxford University Press.
- Kuhn, T. (1970). *The structure of scientific revolutions* (2nd rev ed.). The University of Chicago Press. (Original work published 1962)
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. L. & A. Musgrave (Ed.), *Criticism and the growth of knowledge* (pp. 91–195). Cambridge University Press.
- Latour, B. (1987). *Science in Action*. Harvard University Press.
- Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science*, 48(1), 19–49. <https://doi.org/10.1086/288975>
- Laudan, L. (1996). *Beyond positivism and relativism*. Westview Press.
- Nickles, T. (2019). Historicist theories of scientific rationality. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Summer 2019).  
<https://plato.stanford.edu/entries/rationality-historicist/>
- Popper, K. (1959). *The logic of scientific discovery*.
- Psillos, S. (1999). *Scientific realism*. Routledge. <https://doi.org/10.4324/9780203979648>
- Quine, W. V. (1969). Epistemology naturalized. In *Ontological relativity and other essays* (pp. 69–90). Columbia University Press. <https://doi.org/10.7312/quin92204-004>
- Sokal, A. (1996). Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity. *Social Text*, 14(46/47), 217. <https://doi.org/10.2307/466856>
- Sokal, A., Bricmont, J., & Mermin, N. (1998). Fashionable nonsense: Postmodern intellectuals' abuse of science. *Physics Today*, 52(4), 70–72.  
<https://doi.org/10.1063/1.882506>

- Swift, A. (2017, May 22). In U.S., Belief in Creationist View of Humans at New Low. *Gallup. Com.*  
<https://news.gallup.com/poll/210956/belief-creationist-view-humans-new-low.aspx>
- Trigg, R. (2001). *Understanding social science: A philosophical introduction to the social sciences* (2nd ed). Blackwell Publishers. (Original work published 1985)

THOMAS KUHN



## THOMAS S. KUHN (1962/1970) *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS*

---

Thomas Kuhn's 1962 bestseller *The Structure of Scientific Revolutions* is arguably one of the most influential, but also controversial, works in twentieth century philosophy of science. Kuhn argues that instead of showing cumulative and linear progress, the development of science is rather characterized by sudden revolutionary shifts whereby one 'paradigm' (scientific world picture) is substituted by another. These revolutions are so radical that subsequent paradigms become 'incommensurable', i.e., they cannot be meaningfully compared because there is no common measure. Furthermore, Kuhn argues that a scientific revolution resulting in a paradigm shift is not governed by logical and rational rules but by extra-scientific factors. Kuhn was one of the first thinkers to focus on the socio-historical dimension of science. Yet Kuhn was also criticized for being too pessimistic about the apparent rationality and continuity in science. Some critics even branded him as a radical relativist. In 1970 Kuhn added a Postscript to the second edition of his book in which he tries to refute these accusations. There Kuhn insists that he is not a relativist. He writes: "Later scientific theories are better than earlier ones for solving puzzles [...] That is not a relativist's position, and it displays the sense in which I am a convinced believer in scientific progress."

---



Sonja Smets and  
Fernando R. Velázquez Quesada

**The Study of Informational  
Processes**

# THE STUDY OF INFORMATIONAL PROCESSES

## *An Introduction to Logic*

Sonja Smets and Fernando R. Velázquez Quesada  
[s.j.l.smets@uva.nl](mailto:s.j.l.smets@uva.nl)  
[F.R.VelazquezQuesada@uva.nl](mailto:F.R.VelazquezQuesada@uva.nl)

### 1. INTRODUCTION

Suppose your teacher tells you ‘if you do study hard, then you pass the exam’. If you consider the teacher’s statement to be true and conclude that ‘if you do not pass the exam, then you do not study hard’, you have made a valid inference. However classical logic will tell you that it wouldn’t be a valid inference if you concluded instead that ‘if you pass the exam, then you do study hard’. Indeed, logic is typically understood as the study of *inference*, the process of deriving a conclusion from a set of premises. Even more: for some, logic is actually only about *valid* inferences: those in which the truth of the premises guarantee the truth of the conclusion. The study of valid inference has been one of the main focuses within logic, and in fact it may have been its original motivation. But, through its contact with other domains in philosophy and with different fields, such as economics, computer science and artificial intelligence, logic has become so much more. Modern logic can be understood as the study of **(i)** the different information-changing *actions* that real and/or artificial *agents* (e.g., human beings, but also computer programs) can perform, **(ii)** the way in which these actions *affect* the agents’ individual and collective *information*, and **(iii)** the way in which these actions *interact* with one another, creating long-term information-changing procedures. The aim of this chapter is to provide a brief storyline describing how logic has evolved from focusing on individual human reasoning patterns, such as deciding whether proposition “*q*”

follows from both “ $p$ ” and “ $p$  implies  $q$ ”, to analyzing the flow of information in complex social scenarios where human and artificial agents interact.

## 2. PATTERNS OF REASONING

**A**ristotle (384-322 BC) is considered to be the father of Western logic, his theory of syllogisms has had an unparalleled influence on the history of Western thought. In this theory, Aristotle focused on what he called *deduction*, a concept that nowadays is also called valid inference: a form of inference in which the conclusion follows *necessarily* from the premises.<sup>1</sup> His goal was not to study the properties of inferences used by regular people (what one might call a *descriptive* account); it was rather to identify inferences that people *should* use (what one might call a *normative* account).

One of the features that made the theory of syllogisms so significant is that, when identifying valid inferences, Aristotle focused not on the actual content of the involved statements; rather, he focused on the inference’s *form*. For example, when considering an inference such as “Every Greek is human; every human is mortal. Therefore, every Greek is mortal”, Aristotle did not discuss the meaning of the terms “Greek”, “human”, and “mortal”. He rather directed his attention towards the more general form “Every  $X$  is  $Y$ ; every  $Y$  is  $Z$ . Therefore, every  $X$  is  $Z$ ”, then observing that *any* inference following this scheme is valid, regardless of what  $X$ ,  $Y$  and  $Z$  are. This was a simple but powerful idea, as it allowed his findings to be suitable for all areas of inquiry. Indeed, Aristotle’s framework “... systematizes the principles licensing acceptable inference, and helps to highlight at an abstract level seductive patterns of incorrect inference to be avoided by anyone with a primary interest in truth” (Shields 2016).

Aristotle’s study of syllogisms was devoted to inferences that have two *premises*, three *properties* (the  $X$ ,  $Y$  and  $Z$  of before), and in which premises and conclusion are all statements of the form “(it is not the case that) every/some  $X$  is (not)  $Y$ ”. Yet, the crucial insight of looking for valid *patterns* of inference is probably its most important legacy. From there, several lines of work followed, extending the idea by studying patterns that are not restricted to the ‘simple’ forms of reasoning a syllogism allows. For example, one might reason from more than two premises. Similarly, the premises and the conclusion might involve more than three properties (or maybe less). More interestingly, one might use *connectives*,

---

1. More precisely, a form of inference in which the conclusion is true *in every situation* in which all the premises are true. Equivalently, a form of inference in which *it is not possible* for all the premises to be true and yet for the conclusion to be false. Note thus how, in a valid inference, the premises *do not* need to be true: it is only required that, *if* all the premises are true, then the conclusion *must* be true.

such as *conjunctions* (“and”), *disjunctions* (“or”) and *negations* (“not”), to create more complex statements. Also interesting would be to require a more complex *quantification* pattern than the one present in syllogisms. For this, note how every statement in a syllogism uses exactly one *quantification* over the objects in the domain of discourse, being either a *universal quantifier* (“every”) or an *existential quantifier* (“some”). However, certain areas of inquiry may require more complex combinations, as in case when one wonders whether “somebody is loved by everybody” guarantees that “everybody loves somebody”.

Among the several directions in which the *language* used in syllogisms can be extended, two of them are arguably the most important. The first one, called nowadays *propositional logic*, focuses on the valid reasoning patterns that arise when the premises and the conclusion are built from basic *propositions* (statements that can be qualified as true or false) by means of *connectives* (the already mentioned conjunction, disjunction and negation, and also others, such as implication). For example, within propositional logic one can prove that the famous *modus ponens* (if both “*p*” and “*p* implies *q*” hold, then so does “*q*”) is a valid form of inference. The second one, called nowadays *predicate logic*, focuses on the valid reasoning patterns that arise when the premises and the conclusion are built from basic *predicates* (statements whose truth-value depend on the properties of and relations between the *objects* in the domain of discourse) by means of the mentioned connectives and also arbitrary patterns of universal and existential *quantification* over objects.<sup>2</sup> For example, within predicate logic one can prove that the following inference is valid: if “somebody is loved by everybody”, then surely “everybody loves somebody”. Sharing the use and interpretation of connectives (also called *Boolean connectives*), both propositional logic and predicate logic coincide in several aspects.<sup>3</sup> Together, they constitute the core of what is typically called *classical logic*.

Even though classical logic is based on a number of strong principles, these principles may not always seem so natural when applied to different reasoning scenarios. Hence, some of the essential assumptions and principles of classical logic can be challenged; this gives rise to a variety of systems commonly referred to as *non-classical logics*. For example, classical logics rely on the idea that every sentence is either true or false<sup>4</sup>, but one can allow them to have other

- 
2. Predicate logic’s full name is *first-order* predicate logic, emphasizing the fact that, while quantifiers can act over objects (“all objects have the property *P*”), they cannot act over sets of objects (“all sets of objects have the property *P*”). Quantification over sets of objects is studied in *second-order* predicate logic and *higher-order* predicate logic (the latter includes quantification over objects, over sets of objects, over sets of sets of objects, and so on).
  3. In fact, predicate logic can be seen as an *extension* of propositional logic.
  4. Sentences of the form “*p* or not *p*” and “if both *p* and not *p* hold, then *q* holds” are always true in classical logic, regardless of what *p* and *q* represent. Yet, these principles are dropped,

truth-values<sup>5</sup>. A further departure from classical logic encompasses the systems falling under the umbrella of *modal logic*, in which we allow different modes of truth such as necessarily true or possibly true. These systems are extensions of propositional logic that use additional connectives called *modal operators* (e.g., the *unary operator* “ $\Box$ ”). Depending on their particular interpretation, expressions of the form “ $\Box p$ ” can be read, e.g., as “ $p$  is necessary” (*alethic modal logic*), “ $p$  is obligatory” (*deontic logic*) or “from now on,  $p$  will always be the case” (*temporal logic*). In the context of an agent and its information, two crucial readings are “ $p$  is believed” (*doxastic logic*) and “ $p$  is known” (*epistemic logic*). One of the reasons for the importance of (basic) modal logics is that they can be seen as *classical logics* located between propositional logic and predicate logic.

The identification of valid patterns of reasoning has been a fundamental tool in sciences with a primary interest in filling in the ‘logical consequences’ of what we already know. Hence the importance of logic in, e.g., mathematics and computer science. Nevertheless, the view that logic should be only about one notion of inference (let alone *valid* inference) is rather restrictive and is not in line with the work of the great pioneer Bernard Bolzano (1781-1848), who thought that logic should rather chart the many different consequence relations that people might use, depending on the reasoning task at hand. A similar rich view is found in the work of other philosophers, as John Stuart Mill (1806-1873) and, especially, C. S. Peirce (1839-1914). This interest in a wider variety of inference patterns was encouraged in the early 1980s, mainly through the emergence of artificial intelligence and its search for reasoning patterns that are similar to how humans solve problems. Inferences such as “the grass is wet; therefore, it rained” (*abductive reasoning*, a form of inference that looks for an explanation of an unpredicted observation) and “Pingu is a bird; therefore, it flies” (*default reasoning*, a form of inference usually used in situations with incomplete information) have become the subject of study in modern logic. Such studies are linked to investigations in the empirical sciences where we look for explanations or reason about what is typically (but not always) the case. Note how both inferences are non-valid in classical logic: one can easily think of a situation in which the grass is wet and yet it did not rain (e.g., “the sprinkler is on”), and also of a situation in which Pingu is a bird and yet it does not fly (e.g., “Pingu is a penguin”). In other words, both forms of inference are non-valid because the conclusion might be dropped in the light of further information.

---

respectively, in *intuitionistic logic* and in *paraconsistent logic*.

5. The most popular of such logics include not only the systems in which sentences can have one of *three* possible truth-values (*three-valued logics*), as initially developed by the logician Jan Łukasiewicz (1878-1956), but also those in which the truth-value of a sentence can be any real number between 0 and 1 (*fuzzy logic*).

By looking at forms of reasoning whose conclusions might be dropped when further information arrives (*non-monotonic* reasoning), logic has also expanded its original goal. Its focus is no longer only normative, looking for the reasoning patterns people *should* use. Logicians nowadays are also interested in representing and understanding the many other ‘fallible’ forms of inference that ‘real’ non-omniscient humans use every day. Note that this does not undermine the search for valid patterns; it rather *enriches it*. In a classical experiment, the Wason selection task, subjects are presented with four cards placed on a table, with each card having a letter on one side and a number on the other. The visible faces of the cards show A, K, 4 and 7, and the subjects are asked to point to the card(s) that must be turned over to test whether the rule “if there is a vowel on one side, there is an even number on the other” holds. Most subjects point (correctly) to the card showing A (if the rule holds, its other side should have an even number). However, while most subjects also point (incorrectly) to 4 (the rule does not enforce anything when a side has an even number), only few of them point (correctly) to 7 (if the rule holds, its other side should *not* have a vowel). These results are surprising and have given rise to a series of replication studies and variations of the selection task. Yet, the conclusion from these studies should not be that humans fail to reason ‘correctly’ or that real-life reasoning is very different from what our theories of logic say.

First, when subjects are explained the logical solution of the Wason selection task, they understand it and agree with it. Second, subjects presented with versions of the task based on ‘more familiar’ rules (e.g., “if you drink alcohol here, you have to be over 18”) tend to point to the correct cards. In order to explain this, it has been suggested that the *representation* of the reasoning task plays an important role, and one can also look for the governing principles.<sup>6</sup> George Boole (1815-1864), often considered the father of the purely mathematical approach to propositional logic, regarded the rules of propositional calculus as describing essential human thought. Nevertheless, he acknowledged that human reasoning often deviates from this canon, taking this as a sign that there are further laws that still need to be discovered.

Recognising the important role played by non-monotonic reasoning allows logic to look at ‘more real’ agents. Still, our story so far only focuses on the inferences a single agent can obtain on its own. As such, it highlights a single person’s intellectual effort, as paradigmatically exemplified by Auguste Rodin’s sculpture *The Thinker*. However, as the next section explains, individual inferences are not the only meaningful informational action.

---

6. One explanation that has been offered is that this scenario points to the presence of a classical framing effect, a cognitive bias studied in psychology.

### 3. REASONING IN INTERACTION

Consider the following situation:

“You are in a restaurant with your parents, and you have ordered fish, meat, and vegetarian, for you, your father and your mother, respectively. A new waiter comes with the three dishes. What can he do to know which dish corresponds to which person?”

The waiter can ask “Who has the fish?”; then he can ask “Who has the meat?”. Now he does not have to ask anymore: as the logician Johan van Benthem [b.1949] points out, “two questions plus one inference are all that is needed”.

The example shows how the information flows when agents start *interacting* with their environment. In the example, the answers to the posed questions provide the waiter with information that he otherwise would have not been able to obtain. Maybe more importantly, he can make use of the information he has received to perform further inferences that, together with his initial information (each person gets only one dish, and each dish corresponds to one person), allows him to deduce who gets the vegetarian dish. In other words, it is the *interplay* of these processes, the ‘external’ asking-questions-then-receiving-answers and the ‘internal’ inference, that allows him to get to know what the real situation is.

Despite the fact that logic has looked into informational acts other than inference only relatively recently, the realization that there are further informational actions that should be taken into account was already present in earlier days. In fact, it was already discussed by Aristotle. At the time, it was argued that knowledge could only arise from valid reasoning based on known premises. But if that is true, the only way in which the premises of a valid reasoning can be known is if each one of them is, in turn, the conclusion of some other valid inference. This argument can be used again, now on the newly found inferences, and its repetition becomes then a process that tries to find the ‘original’ premises from which everything else follows. But then, there is a problem. On the one hand, if the process never ends, then nothing can be demonstrated. On the other hand, if the process ends, then these ‘original’ premises are not properly justified, and consequently neither are any of the statements deduced from them. For solving this *regress problem*, Aristotle argues [*Posterior Analytics* I.2, II.19] that the idea that scientific knowledge is only possible by using valid reasoning with scientifically known premises is incorrect: there is another form of knowledge for ‘the original’ premises, and this provides the starting point for demonstrations. Aristotle compares the capacity of having this knowledge with the capacity for sense-perception, which can be understood as the capacity of receiving information from *external* sources. In this sense, it can then be said

that information (or, in the central terms of epistemology, knowledge and beliefs) arise not only from inferences, but also from the interaction with our environment.

With this idea in mind, one can provide a more precise description of the way the information of the waiter (his information state) changes through his interaction with the family. One possible way of describing this process, useful later, is the following. At the start, the new waiter has full uncertainty about who ordered what, yet he knows that three people each ordered a different dish (fish, meat or vegetarian). This leaves open six possible ways in which the world could be (all possible ways in which three dishes can be distributed over three people), with only one of them having the correct person-dish distribution. Without any further information, our waiter cannot distinguish between these six 'possible worlds'. It is only after the first question-answer, when our waiter learns who ordered fish, that he can discard some of these alternatives (in fact, four of them). When he receives the answer to his second question, he can discard one of the two remaining possibilities, thus gaining full knowledge of the correct person-dish allocation.

The waiter thus gets his information from external sources in the form of answers to his questions. As we saw, Aristotle acknowledged the importance of external sources by referring to sense perception. Use of the senses, however, is a limited form of interaction. Sense-perception indeed allows us to get information from external sources, but these external sources are then being understood as 'unresponsive' objects. The informational acts that a human can carry out with these objects can be called *observations*; as useful as they are, they are not the only way in which one can interact with the environment.

In the restaurant example, the new waiter does not have an interaction with objects. He has an interaction with other agents, who might indeed provide him with further information, but who (just as himself) have also knowledge, beliefs, preferences and goals, and therefore also *react* and *change*. In these cases, this exchange can now be properly called *interaction*, and it goes beyond what has been described above. When the waiter asks a question, he reveals the family not only that he does not know the answer, but also that he wants to know the answer, and that he considers it possible for the family to know it. Then, in turn, the family answers because they want to start their dinner, and they know their answer will allow the waiter to know who gets each dish.

It is important to notice that, during the conversation, the new waiter is not the only one whose information changes: the information of everyone else present also changes. If the answers given in reply to the waiter's questions were truthful and truly public (everybody witnesses the action, everybody sees everybody witnessing the action, and so on), then the person-dish distribution has become *common knowledge* among everybody present: everybody knows who gets each



dish (in particular, the waiter received the needed information), everybody knows that everybody knows (in particular, the answers of the family were public, so they know the waiter knows), everybody knows that everybody knows that everybody knows, and so on.<sup>7</sup> Thus, they all get to know something new; in particular, they all get *higher-order* knowledge, that is, knowledge about the knowledge of the agents involved (including themselves). It is worthwhile to notice how common knowledge would not have been reached if the waiter's questions were answered only by one member of the family while the others were not looking or listening: the inattentive agents would not know that the waiter knows.

It is also worthwhile to notice that, while in this scenario the interaction is successful (the important bits of information become common knowledge), the outcome could have easily been otherwise if the agents behaved slightly differently. For instance, if the waiter had other goals, he might have attempted to mislead by asking insincere questions. Likewise, if the members of the family wanted to confuse the waiter, their answer could have been untruthful, or they could have not answered at all. These possibilities are what distinguish 'simple' acts of observation, with one side being an unresponsive object (as in sense perception), from real acts of interaction, with both sides being agents with knowledge, beliefs, preferences and goals.

#### 4. MODELLING INFORMATION AND INFORMATION CHANGE

If the main focus of logic is the way different information-changing actions affect the information of agents, then its formal study requires at least two main ingredients. We need a formal representation of the information that the agents have, and we need a formal representation of the effects of these actions. For the first, the above-mentioned *epistemic logic* (see, e.g., Fagin et al. 1995, van Ditmarsch et al. 2015) is an important tool, as it allows us to represent and reason about the knowledge (or, when understood in a broader sense, the different kinds of knowledge and the different forms of belief) of both individual agents and groups of them. For the second, the ideas behind *dynamic epistemic logic* (see, e.g., van Benthem 2011) are insightful, as they allow us to describe changes in the structures representing the agents' information. In fact, the basics of both epistemic logic and dynamic epistemic logic have been already

---

7. Public actions are the paradigmatic way of reaching common knowledge. For example, when cash is used to pay for a product or a service, both the client and the cashier tend to put the money (the payment and the change, respectively) in a tray. In this way, both have common knowledge about the amount of money being exchanged. Similarly, when applying an injection, the specialist will unwrap the syringe and needle in public view, so it will be common knowledge that they are brand new.

described, albeit implicitly, when analysing the restaurant example. Here it is a more precise presentation of these ideas.

To represent the knowledge of an agent, we adopt the basic idea of *information as range*. An agent's available information is essentially represented as a range (collection) containing all the different ways the world could be from her perspective (i.e., all worlds the agent considers possible). In settings that follow this idea, the *knowledge* of an agent is defined as what holds in all those alternatives, that is,

an agent knows that a certain proposition is true  
whenever the proposition is true in *all* worlds/situations in her range.

Note how, in this approach, the model does not represent directly the knowledge of the agent: it rather represents the agent's *uncertainty*, and then defines her knowledge in terms of it. In the restaurant example, the initial information of the new waiter can be modelled by a range containing the six possible ways in which the three dishes can be distributed over the three persons.

The discussion above suggests assigning a set of 'possible worlds' to each agent. Yet, it is more practical to represent the range by means of a relation connecting worlds to worlds. Thus, given a possible world, the agent's *indistinguishability* relation indicates which are the worlds that the agent considers possible from it. These structures for representing an agent's knowledge, called *possible worlds model*, *Kripke models* (after the philosopher Saul Kripke [b.1940]) or, more generically, *relational models*, have several advantages. An important one is that the move from a single-agent case to a multi-agent scenario becomes straightforward: we only need a separated depiction of the worlds each agent considers possible, and this can be done by means of different indistinguishability relations. A more interesting advantage is that, thanks to the relational representation, these structures describe not only what an agent knows about the real world, but also what they know about what *they and other agents* know about the real world, and so on. As discussed above, this form of *higher-order* knowledge is crucial when an agent interacts with others.

It is important to notice that the notion of knowledge depicted by these structures is very strong. Before considering an example, it is worthwhile to briefly mention some of the features of this form of knowledge. The most noticeable characteristic is that knowledge is 'closed' under modus ponens. Indeed, if an agent knows both "*p*" and "*p* implies *q*", then both are true in every world in the agent's range; but then, "*q*" is also true in each one of those worlds, and hence it is known by the agent. Thus, agents whose knowledge is represented in this way do not need to perform modus ponens: they already have everything they

would get from it.<sup>8</sup> Another important feature of relational models refers to the consequence of making the assumption that every world is indistinguishable from itself, which implies that the agent always includes the real world in her range. Under this natural and seemingly harmless postulate, whatever is known by the agent *needs to hold in the real world*, and therefore cannot be revised (this is called among philosophers *irrevocable* knowledge). Finally, under other reasonable assumptions about the indistinguishability relation (and hence the range), the knowledge of an agent is both *positively and negatively introspective*. This means that, if the agent knows that something is true, she knows that she knows it (positive introspection) and if she does not know that something is true, she knows that she does not know it (negative introspection). All of these features can and have been debated extensively, with epistemologists proposing additional fine-grained epistemic attitudes (e.g., explicit knowledge, defeasible and/or non-introspective knowledge, different forms of beliefs) and logicians proposing different structures to represent them (e.g., plausibility models, neighbourhood models, evidence models). Yet, the strong form of knowledge represented by relational models is enough to exemplify our main point: agents have knowledge, and their knowledge can change when they communicate with one another.

Since an agent's knowledge is defined by its range, actions that change her knowledge can be represented by operations that change her range. In particular, consider the action that provides the agents with truthful information via public communication. By stating that a given proposition is true, the action is effectively ruling out all those situations (in this case, possible worlds) in which the proposition is false. Thus, an act of public observation can be represented by an operation that *reduces* the agents' range. Note how this works as one intuitively expects: by eliminating worlds, the operation decreases the agent's uncertainty, thus increasing her knowledge. In the restaurant example, this is exactly the operation used for representing the way the answers of the family affect the new waiter's information state.

Here, a word of caution is needed. A range-reducing operation is adequate for representing acts of *truthful* and *public* communication. But, if the communication were public and yet not truthful, then worlds could not be eliminated; they would need to be somehow 'downgraded', so they do not play a role when defining the information the agent is truly entertaining, and yet they remain available in case the provided information is proven to be false. In fact, if the communication is not truthful, one might rather stop calling the agent's information *knowledge*, and start calling it *belief*. Likewise, if the communication were truthful and yet

---

8. This also explains the mismatch between the initial analysis of the restaurant example (which indicated that two questions plus one deductive inference were needed) and the more detailed description provided later (in which the waiter only needed the answers to his two questions).

not public, again worlds could not be eliminated, but now for a different reason. The agents receiving the information certainly should not take those worlds into account anymore when looking for what they know about the real world; yet, the agents that do not receive the information certainly cannot discard them. More importantly, even agents that received the information need those worlds: they need them to evaluate what they know *about what is known by agents that did not received the information*. But once again, despite these caveats, the range-reducing representation is enough to make our point (knowledge changes), and in fact also enough to find subtle details about the way information flows.

To explain the mechanism of knowledge updates and information change in more detail, here is another example: the high-school Math Olympiad puzzle called 'Cheryl's birthday', which was part of the Singapore and Asian Schools Math Olympiad 2015 Contests. In it, several agents are given both private and public information; in order to solve the problem, they have to reason about both the facts at hand and the information states of others. The puzzle reads as follows:

Albert and Bernard have just become friends with Cheryl, and they want to know when her birthday is. Cheryl gives them a list of 10 possible dates:

- May 15, May 16, May 19,
- June 17, June 18,
- July 14, July 16,
- August 14, August 15, August 17.

Cheryl then tells Albert the month of her birthday but not the day, whereas she tells Bernard the day but not the month. Afterwards, the following conversation takes place:

- *Albert*: I don't know when Cheryl's birthday is, but I know Bernard does not know either.
- *Bernard*: At first I didn't know when Cheryl's birthday is, but I know it now.
- *Albert*: Then I also know when Cheryl's birthday is.

The question to the reader is: *When is Cheryl's birthday?*

The correct answer is July 16th. But how can we be so sure about it? In order to solve the puzzle, we have to reason about the information state of each of the agents. In other words, we need to ask: what does each agent know? However, just representing the agents' knowledge will not do: we also have to reason about the effect that the acts of communication have on their information state.

In the analysis below, we make use of the above described tools of epistemic and dynamic epistemic logic.<sup>9</sup>

To start, one first needs to analyse who knows what exactly before the conversation between Albert and Bernard takes place; this will be done using a possible worlds model, similar to the way we described the different possible worlds in the restaurant example. Cheryl has placed 10 options on the table, so we start with 10 possible worlds (labeled from  $s_1$  to  $s_{10}$ ), each one of them representing a possible date for Cheryl's birthday: in  $s_1$  her birthday is on May 15th, in  $s_2$  it is on May 16th, in  $s_3$  it is on May 19th and so on. With the possible worlds provided, it is possible to indicate Albert and Bernard's (different) ranges by providing their respective indistinguishability relations. For instance, Albert knows the month of Cheryl's birthday. Thus, he can distinguish any 'May'-world from any 'non May'-world; likewise for any other month. However, he does not know the day of Cheryl's birthday. From his perspective, all 'May'- worlds ( $s_1$ ,  $s_2$  and  $s_3$ ) are indistinguishable from one another; likewise for any other month.

Let us now see what happens to our model when the above described conversation between Albert and Bernard takes place. Albert's first statement reveals his own uncertainty about Cheryl's birthday but adds a very important piece of information: he knows that Bernard does not know it either. Recall that Bernard knows the day (but not the full date) of Cheryl's birthday. Hence, if the day were 18th or 19th, he would immediately know the month too (and hence the full date); this is because each of these numbers is true in only one possible world, and thus Bernard would know immediately which the real world is. Thus, for Albert to know that *Bernard doesn't know*, it has to be so that, in *all* worlds that Albert considers possible, Bernard doesn't know the date.

Now, in case Albert would have been told the month was May, then he would not have known that Bernard doesn't know the date: in May, there is a possible world in which Bernard *knows* (19th). The same reasoning holds for June, as it contains a day (18th) in which Bernard knows the full date. But Albert does indeed know that Bernard doesn't know the exact date, so Cheryl's birthday cannot be in May, and neither in June. Thus, when Albert announces (we assume truthfully, and definitively publicly) that he knows Bernard does not know, he is effectively announcing that the birthday cannot be in May or June. Thus, both Albert and Bernard have now only five possible worlds left to reason about: the possible dates in July and August.

---

9. One can graphically represent both Albert and Bernard uncertainty at each stage of the conversation about Cheryl's birthday. Section 1 of Appendix B in Baltag and Renne (2016) provides us with the required drawings (Figures B1-B4) that will help the reader to follow the different reasoning steps that we describe in this section.

The conversation is continued by Bernard, who announces publicly that *now* he knows Cheryl's birthday. If this is a truthful statement, then the day cannot be the 14th (as Bernard would not know the full date: he would not know whether the month is July or August). No other options can be discarded: for any of the other days that are still left, if that is the true one, then Bernard knows it. This leaves us with a model with only three possible worlds left.

In the final communication step, Albert says that *now* he also knows when Cheryl's birthday is. In the current model with three possible worlds, only in one of those possible worlds it is true that Albert knows the full date: the 16th of July. Thus, under the assumed truthfulness of all involved parties, this is which Cheryl's birthday is.

Note how this analysis of the birthday scenario depends on a number of assumptions which are crucial to reach the conclusion we have drawn. For instance, again, it is implicitly assumed that the agent's information refers to their *irrevocable knowledge* (that is, knowledge that cannot be revised). It is also assumed that Cheryl, Albert and Bernard have *common knowledge* about a number of basic facts. It is common knowledge that Albert and Bernard did not know Cheryl's birthday *before* Cheryl handed them the 10 possible options; it is also common knowledge that both Albert and Bernard know that the correct answer is among the 10 given options. Further, it is also common knowledge that, immediately before his conversation with Bernard, Albert knows the precise month, but nevertheless does not have information about the day other than what is implied by the 10 listed dates. In other words, at this stage Albert knows the month, but not the day. Analogously, what Bernard knows exactly before his conversation with Albert is commonly known too. Moreover, it is assumed that Albert and Bernard tell the truth when they speak and are fully heard and understood in their public conversation. Finally, it is also assumed that Albert and Bernard have immediate access to every deductive consequence of what they know. Any deviation of these assumptions may lead to a different outcome of the problem.

Yet, the formal tools used through the analysis allow us to look at subtle details that otherwise could not have been noticed. One of them is relative to notions of knowledge for *groups* of agents. The notion of common knowledge has been already mentioned, but there are other important group epistemic attitudes. One of them is that of *distributed knowledge*: the knowledge obtained when a group of agents put together all they individually know and include all its logical consequences (i.e., everything that can be *deduced* from it). In the example, Cheryl's birthday is distributed knowledge among Albert and Bernard once Cheryl has told Albert and Bernard, separately and respectively, the month and the day. Are there other relevant group attitudes? Concepts like these are important when modelling, e.g., distributed systems within computer science.

Note also how, from his perspective, Albert and Bernard's conversation can be seen as a successful communication process: it transforms their distributed knowledge of the full date of Cheryl's birthday into common knowledge. But then, one can wonder: what if Bernard had talked first? Or, what if, instead of taking turns, only one of them was allowed to talk? Or, what if there were restrictions on what they could communicate? In general, one can wonder the following: what would have been the outcome of the interaction if the conversation had a different structure? This shows the importance of the communication's *protocol*: the established sequence of actions through which the agents will interact. In the birthday example, the protocol that the conversation between Albert and Bernard follows (Albert starts and then they take turns) is successful. Are there other protocols that are successful in this scenario? Would this protocol be successful in a different situation? Questions like this are relevant in computer science, in particular in communication within computer networks.<sup>10</sup>

If one is interested in this kind of questions, one might be also interested in a more fundamental issue. In the example, the initial information that Cheryl provided, together with the conversation Albert and Bernard had, allowed them to get to (commonly) know a specific truth: Cheryl's birthday is on July 16th. But, what about other facts? Is their initial information and communication protocol enough for them to get to know *everything* that is relevant to the situation? More generally, and taking a philosophical perspective, can an agent get to know all true sentences? In philosophy, advocates of the *verificationist thesis* answer this question positive, and yet opposite arguments as Frederic Fitch's *paradox of knowability* have been proposed. In the epistemic logic setting, it turns out that not all truths can be known. An agent might not know that a certain fact "*p*" is true, thus making the sentence "*p* is true and she does not know it" also true. However, there is no way for her to (irrevocably) know that "*p* is true and she does not know it". If she were told that the sentence is true, then she would learn its first part: *p* holds. But then, she cannot know its second part, that she herself does not know *p*: now she knows it. This shows how these formal tools can shed new light on philosophical issues.

## 5. FURTHER PERSPECTIVES

**W**hile the restaurant example shows how informational actions are not restricted to acts of inference, the analysis of the birthday example sketches how 'complex' real-life scenarios of multi-agent interaction

---

10. For example, the *Russian card problem*, discussed in cryptography-related circles, poses the following question: can two agents share information in order to make a given secret common knowledge among them, while still making sure that a third agent, eavesdropping on the communication, does not get to know the secret?

can be modelled with logical tools. These examples are meaningful for modern logic, which can be understood as the study of *(i)* the different information-changing *actions* that real and/or artificial *agents* (e.g., human beings, but also computer programs) can perform, *(ii)* the way these actions *affect* the agents' individual and collective *information*, and *(iii)* the way these actions *interact* with one another, creating in this way long-term information-changing procedures.

These examples also show how this broader view on logic establishes meaningful connections with other fields, thus raising further interesting questions about the way information flows in multi-agent systems. First, the discussion on non-valid inference and the attempts to model forms of human reasoning connect logic not only with artificial intelligence, but also with psychology and cognitive science. The latter are particularly useful not only when looking for patterns of inference that are used by humans, but also to explain why other patterns are not. When factors that play a role are identified (e.g., source of information, emotions, moral values), one can incorporate them into a logical formalism, thus providing a more faithful descriptive model of the way humans reason.

Then, the further discussion on the birthday's example highlights some questions that are also important in philosophy and computer science. It has been already argued why individual agents cannot get to know everything that is true (see the end of previous section). But then, can we characterize the kind of truths that an individual agent can get to know? Equally important: how? That is, which are the sequences of actions that lead to this outcome? Of particular importance: can agents get to know what they *distributively* know? If so, can they further *commonly* know everything they distributively know? And, if so, again: how?

These are not the only connections this broader view of logic allows for. As simple as it is, the restaurant example can already be understood as a scenario of *game theory*, a powerful mathematical tool used for analyzing competitive situations in which the outcome of the 'game' (e.g., how many 'points' each agent will get) depends not only on the action an individual agent chooses, but also on the choices of the other agents. Indeed, as discussed, the waiter had several options for her questions to the family (e.g., sincere or insincere), and so did the family for their answers (truthful or deceitful, public or private). They made their choices according to the common goal of having the dishes distributed correctly, but had their goals been changed, their actions might have varied, and the outcome could have been different. In scenarios like this, an information-theoretic and logical perspective has added value. By allowing a representation of the potentially different knowledge, beliefs, preferences and goals the agents might have, it allows us to focus not on what each agent should do given the situation, her preferences and her available actions, but rather on what each one of them can do given her *knowledge* about the situation, the *information* she has



about everybody's preferences, and the actions she *considers* are available (van Benthem et al. 2011, van Benthem 2014).

Another natural connection that can be made is with the social sciences. We provided the reader with different examples which illustrate that understanding the epistemic social environment in which we live and operate is crucial for the success of our individual daily decisions and actions. The examples we gave are not straightforward, as the ongoing discussion on social platforms reveal: a quick and superfluous analysis of the Cheryl's birth puzzle may lead you straight to the wrong answer. Yet, one can learn how to analyze one's own information state as well as the information state of others. Similarly, one can learn how to analyze *the explicit and implicit* information present in the social environment we operate in. In our presentation of the restaurant example and Cheryl's birthday scenario, we did focus mainly on the *explicit* exchange of information: questions are asked and answers are provided, and statements are publicly communicated. But there are also scenarios in which our beliefs, preferences and opinions are affected by more subtle acts, and even only by the (actual, imagined or implied) presence of others. This phenomenon, called *social influence*, can take many forms, as socialization (inheriting and disseminating norms, customs, values and ideologies), conformity (changing attitudes, beliefs and behaviours to match those of the majority), compliance (changing favourably in response to explicit or implicit requests made by others), reactance (adopting a view contrary to what the person is being pressured to accept) and obedience (changing in response to a direct command from an authority figure). This perspective opens up a new line of work in which logic can be used to study informational processes in social systems, including the study of different social epistemic phenomena, such as informational cascades (a number of people make the same decision in a sequential fashion) and pluralistic ignorance ("no one believes, but everyone thinks that everyone believes").

We highlighted the richness of logic and the usefulness of its tools to model different types of scenarios. However, even while the examples in this chapter were all analysed by making use of small-size models of low-complexity, it is important to note that the full power of our tools will become more apparent when the complexity of the models increases by adding agents, propositions, actions and by expanding on the type of attitudes.<sup>11</sup> Overall, when equipped with

---

11. We refer the interested reader to Baltag and Moss (2004) and Baltag et al. (2008) for further information about different such scenarios. For readers who are new to logic, a systematic study will require to get first acquainted with the basics of classical logic, followed by a study of modal logic in order to construct the type of possible worlds models we have informally sketched in this chapter. We refer to van Benthem (2010) for an introduction to the required tools of modal logic, after which the reader can proceed with a study of dynamic epistemic logic. The entry in Baltag and Renne (2016) provides an excellent overview of the state of the

the tool-kit described in this chapter, ranging from classical logic all the way to non-classical logics, the reader will be ready to start navigating the spaces in which different informational processes appear. You will find the logical tools to be a crucial ingredient when operating in a variety of different contexts, including in our digital-social environments which have become increasingly more complex and allow for new forms of information-exchange every day.

---

art on dynamic epistemic logic and supplements the book van Ditmarsch et al. (2015) on this topic.

## BIBLIOGRAPHY

- Baltag, A., & Moss, L. (2004). Logics for epistemic programs. *Synthese*, 139(2), 165–224. <https://doi.org/10.1023/b:synt.0000024912.56773.5e>
- Baltag, A, & Renne, B. (2016). Dynamic epistemic logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016).
- Baltag, A., Van Ditmarsch, H., & Moss, L. (2008). Epistemic Logic and Information Update. In P. Adriaans & J. van Benthem (Eds.), *Philosophy of information* (Vol. 8, pp. 361–455). Elsevier. <https://doi.org/10.1016/b978-0-444-51726-5.50015-7>
- Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. The MIT Press. <https://doi.org/10.7551/mitpress/5803.001.0001>
- Shields, C. (2016). Aristotle. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Stanford University Press.
- van Benthem, J. (2010). *Modal logic for open minds*. CSLI Publications.
- van Benthem, J. (2011). Logical dynamics of information and interaction. In *Logical dynamics of information and interaction*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511974533>
- van Benthem, J., Pacuit, E., & Roy, O. (2014). Toward a theory of play: A logical perspective on games and interaction. *Games*, 2(1), 52–86. <https://doi.org/10.3390/g2010052>
- van Ditmarsch, H., Halpern, J. Y., van der Hoek, W., & Kooi, B. (Eds.). (2015). *Handbook of epistemic logic*. College Publications.

ROSE RAND



## JOHAN VAN BENTHEM

### *LOGICAL DYNAMICS OF INFORMATION AND INTERACTION*

---

**V**an Benthem's "Logical Dynamics of Information and Interaction" (2011) represents a unified account of a wide range of informational processes. Using the framework of dynamic epistemic logic as a vehicle, the book shows how a variety of different insights coming from Computer Science, AI, Philosophy and Cognitive Science can be successfully combined. The book starts by providing formal representations of different epistemic attitudes an agent might have towards information (e.g., several kinds of knowledge, different forms of beliefs) as well as the different actions that can affect them (e.g., the already mentioned inference, observations and different forms of belief revision). From these basics, the book follows different directions that connect logic with concerns in different fields. The study of what these informational actions can achieve connects not only with concerns in Philosophy; but also with issues in Cognitive Science, Computer Science, Artificial Intelligence and Economics. To use Johan van Benthem's own elegant formulation, "my program of Logical Dynamics calls for identification of a wide array of informational processes, and their explicit incorporation into logical theory, not as didactic background stories for the usual concepts and results, but as first-class citizens."

---

Catarina Dutilh Novaes

**'You can't argue  
with that!'**

# 'YOU CAN'T ARGUE WITH THAT!'

## *An Introduction to Argumentation*

Catarina Dutilh Novaes  
[c.dutilhnovaes@vu.nl](mailto:c.dutilhnovaes@vu.nl)

### 1. INTRODUCTION

Suppose that you and a group of friends need to decide what to do together on a Friday evening. There are a couple of alternatives, such as going to a restaurant, to the cinema, going out dancing, or even staying in. Predictably, there is disagreement on what the best option would be: Fatima wants to go to the cinema, Pedro to the restaurant, and Astrid wants to go dancing. But you all agree that doing something together would be more fun than to split up in smaller groups. How do you go about resolving the disagreement?

There are different strategies you may pursue: you may draw a lottery; you may vote without discussing the options beforehand, for example by raising hands; you may engage in negotiations such that the defenders of the losing proposals get something in return for their willingness to concede; one person may decide that she alone should make the call; or you may consider the pros and cons of each proposal, exchanging reasons for or against each of them. If you choose the latter strategy, then you and your friends will be engaging in *argumentation*; the goal would be to converge into a common plan that everyone will feel at least reasonably comfortable with by critically examining the different options together.

Argumentation can be defined as the communicative activity of producing and exchanging reasons in order to support, defend or challenge claims

and positions, especially in situations of doubt or disagreement (Lewinski & Mohammed, 2016). Argumentation is thus best conceived as a kind of dialogue, but it is a *special* kind of dialogue. Indeed, most of the dialogues we engage in are not instances of argumentation. For example when you ask someone if they know what time it is, or when a friend tells you about their vacation, these will typically not be instances of argumentation. Argumentation only kicks in when, upon making a claim, someone receives a request to further justify or provide reasons for that claim, or feels herself that further clarification is required. Since most of what we know we learn from others, argumentation is an important mechanism to filter the information we receive, rather than having to accept what others tell us uncritically.

Thus understood, argumentation is an important but demanding human practice. Across societies and cultures (albeit perhaps to different degrees), argumentation permeates scientific inquiry, legal procedures, education, and political institutions. And indeed, it may even be used to solve mundane problems such as what to do with friends on a Friday evening. The study of argumentation is an interdisciplinary field of inquiry, involving philosophers, communication theorists, legal scholars, cognitive scientists, computer scientists, political scientists, anthropologists, among many others, who address different questions. What is the 'point' of argumentation, i.e. which function(s) does it fulfill? What are the underlying cognitive mechanisms involved? To which extent is persuasion essential for good argumentation? Does argumentation indeed deliver the results we expect from it, such as more accurate beliefs and better-grounded decisions?

A classical approach to argumentation is represented by the pragma-dialectical tradition (Van Eemeren & Grootendorst, 1996) (Van Eemeren & Grootendorst, 1992), and by the informal logic tradition (Groarke, 2016). Authors in these traditions focus on the content and structure of arguments, including the formulation of argument schemata that help analyze and evaluate specific arguments. In this chapter, we take a different approach: rather than looking "inside" arguments, we focus on the uses and goals of argumentation against the background of concrete situations in human life, in particular with respect to the circulation and production of knowledge, social coordination, and whether argumentation can also be used for domination and overpowering.

This chapter starts by presenting three common conceptions: 1) argumentation as an epistemic practice; 2) argumentation as aiming at consensus and social coordination; 3) argumentation as a competition where an arguer's main goal is to "beat the opponent".<sup>1</sup> It then discusses the outcomes of argumentation,

---

1. This tripartite distinction roughly follows the distinction proposed by Lumer (Lumer, 2010) between epistemological, consensualist, and rhetorical approaches to argumentation.



i.e. whether it achieves its presumed goals such as leading to more accurate beliefs or consensus. Finally, we briefly discuss specific institutions where argumentative practices are crucial, namely education, science, the law, and politics.

## 2. ARGUMENTATION AS EPISTEMIC PRACTICE

**W**e speak of argumentation as an epistemic practice when we take its primary purpose to be that of improving our beliefs and increasing knowledge. To engage in argumentation can be a way to acquire more accurate beliefs: by examining critically reasons for and against a given position, we would be able to weed out the weaker, poorly justified beliefs (likely to be false) and end up with stronger, suitably justified beliefs (likely to be true). It is in this sense that argumentation is thought by many people, philosophers in particular, to be *truth-conducive* (Betz, 2013), at least in cases where there is an objectively correct answer to a problem. Epistemologist Alvin Goldman captures this idea in the following terms:

Norms of good argumentation are substantially dedicated to the promotion of truthful speech and the exposure of falsehood, whether intentional or unintentional. [...] Norms of good argumentation are part of a practice to encourage the exchange of truths through sincere, non-negligent, and mutually corrective speech. (Goldman, 1994) (p. 30)

It is at least in theory possible to engage in argumentation with oneself, solitarily weighing the pros and cons of a position. But a number of philosophers, most notably John Stuart Mill, maintain that *interpersonal* argumentation, involving people who truly disagree with each other, best realizes the epistemic goals of improving our beliefs. Mill famously developed this idea in *On Liberty* (1859/1999): when our ideas are challenged by engagement with those who disagree with us, we are forced to consider our own beliefs more thoroughly and critically. The result is that the remaining beliefs, those that have survived critical challenge, will be better grounded than those we held before such encounters. As Mill puts it in *On Liberty*: “Both teachers and learners go to sleep at their post, as soon as there is no enemy in the field.” (Mill, 1859/1999) (p. 83)

Dissenters thus force us to stay epistemically alert instead of becoming too comfortable with existing, entrenched beliefs. But for this process to be successful, dissenters must be permitted to voice their opinions and criticism freely, and indeed Mill’s forceful defense of free speech is one of his most famous positions.

---

However, the rhetorical approach is not accurately captured in the ‘argumentation as power game’ conception to be discussed below.

He emphasizes the role played by the free expression of ideas in facilitating the growth of knowledge in a community: the more dissenting views and arguments in favor or against each of them are exchanged, the more likely it is that the “good” ones will prevail (Halliday & McCabe, 2019). This general idea has been further developed in recent years by philosophers working on the epistemology of democracy, who maintain that one of the strengths of a democratic system where citizens can openly exchange arguments (a deliberative democracy—see next section) is that it leads to better grounded beliefs on political matters (Peter, 2016).

The idea that argumentation can be an epistemically beneficial process is almost as old as philosophy itself. In the three major ancient philosophical traditions—Greek, Indian, Chinese—argumentation is viewed as an essential component of philosophical reflection. In the Greek tradition in particular, *dialectic*, or “the art of conversing,” occupied a prominent position, as registered both in Plato’s dialogues and in Aristotle’s logical texts (Dutilh Novaes, 2021). Dialectic corresponds to a specific form of argumentative interaction, where one participant asks questions in order to get the other participant to commit to specific claims, and then draws conclusions from those commitments, as illustrated in Plato’s dialogues.

In what is often described as the “Socratic method,” in many of Plato’s dialogues Socrates starts by getting his interlocutor A to grant a given thesis, say  $p$ . He then secures further commitments from A, say  $q$  and  $r$ . But he then shows that  $q$  and  $r$  together in fact imply *not- $p$* , thereby showing that A’s collection of commitments  $\{p, q, r\}$  is incoherent, which amounts to a *refutation (elenchus)* of A’s position. A is then compelled to revise her beliefs, as at least one of them has to go. The thought is that, having been thus refuted and so led to revise her beliefs, A will be in a better epistemic position than before. As Socrates puts it when addressing his interlocutor Callicles in Plato’s dialogue *Gorgias*: “Then I’ll be very grateful [...] to you if you refute me and rid me of nonsense; now don’t be slow to benefit a friend, and refute me.” (*Gorgias*, 470c7-10) So even if dialectic cannot establish the absolute truth of a given thesis (“I only know that I know nothing” is a slogan famously attributed to Socrates), it can rid one of nonsense, and that is a desirable epistemic result (Dutilh Novaes, 2020b).

In sum, a number of key figures in the history of philosophy defended the view that argumentation is primarily an epistemic activity which allows knowers to increase their number of true beliefs, decrease their number of false beliefs, and deepen their overall understanding of different issue. Whether it can indeed perform this function will be discussed in section 5 below.

### 3. ARGUMENTATION AIMING AT CONSENSUS AND SOCIAL COORDINATION

Another important strand in the literature on argumentation are theories that view *consensus* as the ultimate goal of an argumentative process: to eliminate or resolve a difference of (expressed) opinion. The influential tradition of pragma-dialectics is perhaps the most prominent recent exponent of this approach (Van Eemeren & Grootendorst, 1992). What seems to motivate these consensus-oriented approaches is the attribution of a role of *social coordination* to argumentation. Because humans are social animals who must often cooperate with other humans to successfully accomplish certain tasks, they must have mechanisms to align their beliefs and intentions, and subsequently their actions (Tomasello, 2014). The thought is that argumentation would be a particularly suitable mechanism for such alignment, as an exchange of reasons would make it more likely that differences of opinion would decrease (Norman, 2016). This may happen precisely because argumentation would be a good way to track truths and avoid falsehoods, as discussed in the previous section. Participants engaging in the same epistemic process of exchanging reasons would all come to converge towards the truth, and thus come to agree with each other. However, consensus-oriented views need not presuppose that argumentation is truth-conducive: the ultimate goal of argumentation on these views is that of social coordination. In short, people can come to agree on certain points without those necessarily being true.

Let us again return to the example of the plans for a group outing on a Friday evening. It is not clear that there should be, objectively speaking, one alternative that is obviously superior to all the others. Arguably, there may well be no truth of the matter when it comes to issues of taste and preferences on how to spend a pleasant evening with friends. Instead, the purpose of collectively deciding what to do and where to go next Friday would be to maximize satisfaction and group cohesion, in particular if those who initially had different preferences come to embrace the option that eventually prevails thanks to the reasons offered by others (thus coming along happily rather than grudgingly). On this picture, it is important that everyone feels 'heard' during the deliberative process so that they perceive the final decision as the result of a truly collective process; the procedure matters just as much as the result itself.

Indeed, the social complexity of human life is ultimately what is behind the emphasis on consensus. The Friday evening example is in fact a fairly simple situation when compared to a number of other important situations where some degree of consensus and coordination is necessary, especially *political decisions*. In political theory, the concept of *deliberative democracy* rests crucially on argumentative practices (Landemore, 2012) (Fishkin, 2016). (For

present purposes, “deliberation” and “argumentation” can be treated as roughly synonymous terms). In a deliberative democracy, for a decision to be legitimate, it must be preceded by authentic public deliberation—a discussion of the pros and cons of the different options—not merely the aggregation of preferences that occurs in voting. Even if democratic deliberation does not lead to full consensus, the different people involved may opt for a compromise solution. This is what usually happens in, for example, coalition-based political systems (such as in the Netherlands), where after an election typically a number of different parties must come together in a coalition to compose a majority government.

A prominent theorist of deliberative democracy thus understood is Jürgen Habermas, whose “discourse theory of law and democracy” relies heavily on practices of political justification and argumentation taking place in what he calls “the public sphere” (Habermas, 1996). He starts from the idea that politics allows for the collective organization of people’s lives, including the common rules they will live by. Political argumentation is a form of communicative practice, so general assumptions for communicative practices in general apply. However, additional assumptions apply as well. In particular, deliberating participants must accept that anyone can participate in these discursive practices (democratic deliberation should be inclusive), and that anyone can introduce and challenge claims that are made in the public sphere (democratic deliberation should be free). They must also see one another as having equal status, at least for the purposes of deliberation (democratic deliberation should be equal). (Olson, 2014)

Habermas’s discourse theory of democracy thus presupposes a fair amount of common ground among interlocutors for deliberative processes to take place legitimately. It also requires equal status for those involved. Critics of Habermas’s account view it as unrealistic, as it presupposes an ideal situation where all citizens are treated equally and engage in public debates in good faith (Mouffe, 1999) (Geuss, 2019). We all know that, in practice, in private as well as public debates, people engage in all kinds of vicious behavior when arguing. Indeed, the conception of argumentation to be discussed in the next section highlights precisely the adversarial aspects of argumentative practices (in contrast to the cooperative, consensual aspects discussed until now).

#### 4. ARGUMENTATION AS A POWER GAME

The two conceptions of argumentation discussed so far offer a largely optimistic picture of the outcomes of an argumentative process: it may lead to more accurate beliefs, or it may lead to consensus and to social coordination. But argumentation also seems to have a “dark side,” with its potential to be used for coercion and manipulation. In fact, argumentative encounters often seem to turn hostile, so much so that, in English, “to have an argument with someone” means the same as to have a quarrel or a fight. In many argumentative situations, participants seem primarily interested in affirming their (intellectual, moral) superiority over others. Even if their motives are not so explicitly related to overpowering others, one might say that any attempt to persuade someone of a view that they do not initially endorse is in fact a form of coercion (Nozick, 1981). On this view, rather than a cooperative endeavor to seek truth or reach consensus, argumentation is above all an *adversarial power game*, where arguers seek to dominate or manipulate interlocutors and to score points.

Back to our scenario of friends trying to decide what to do together on a Friday evening, it may well happen that, if they start exchanging reasons for and against the different options, one of the members of the group who happens to be more outspoken and verbally articulate will control the discussion and overpower the others, eventually imposing her favorite option on those whose opinions do not get a fair chance of being heard. Alternatively, an arguer can manipulate an audience by selectively choosing arguments that support her position while concealing arguments against that position.

It is often remarked that argumentative situations *de facto* often escalate towards real fights, especially in specific contexts such as philosophical debates (Moulton, 1983) (Rooney, 2012). It is telling that even the cover of an influential textbook in argumentation theory, *Fundamentals of Argumentation Theory* (Van Eemeren & Grootendorst, 1996), features a scene of two men fighting. Moreover, as noted in (Lakoff & Johnson, 1980) and (Cohen, 1995), the very vocabulary used to refer to argumentative practices is full of references to combat: “He attacked every weak point in my argument.” “His criticisms were right on target.” “Your claims are indefensible.” “I demolished his argument.” “I’ve never won an argument with him.” “If you use that strategy, he’ll wipe you out.” “He shot down all of my arguments.” (Lakoff & Johnson, 1980, p. 4) The pervasiveness of such warlike metaphors in connection with practices of argumentation is a clear indication of the strong association in people’s minds between engaging in argumentation and fighting.

This is not a recent phenomenon; similar complaints were also voiced in ancient Greece. Indeed, being overly adversarial in argumentation is precisely one of the

criticisms addressed at the sophists/rhetoricians by Socrates in Plato's dialogues (e.g. the dialogue *Gorgias*, Gorgias being one of the most famous rhetoricians of his time). According to Socrates (who, as we saw above, takes the goal of argumentation to be that of revealing deeper truths about ourselves and about how to live a virtuous life), the sophists are only interested in overpowering their interlocutors and "scoring points" in purely competitive encounters (Nehamas, 1990).

Admittedly, Socrates is sometimes singularly nasty in these dialogues, but he typically tailors his argumentation to the specific social position, interests, and dispositions of his interlocutors (Moulton, 1983). Arguably, he is nasty only when a more confrontational approach is what is required to deal with recalcitrant interlocutors. The contrast between rhetoricians (as portrayed by Plato at least, which is most likely a somewhat unfair characterization) and philosophers can be captured in terms of the different interpersonal attitudes that Plato attributes to each of them; while the rhetorician only seeks to dominate or win over an audience, the philosopher seeks to benefit others. So for Plato, argumentation understood in the philosophical sense requires a form of *care*, both for oneself and for others. (Irani, 2017)

In recent decades, a number of feminist thinkers have criticized argumentation as an inherently aggressive, adversarial practice (Hundleby, forthcoming). Argumentation would rest on a crystallization of gendered categories of aggressiveness and violence, typically viewed as masculine traits. An implication of the association between argumentation and masculinity is the potential exclusion of feminine gendered persons from argumentative processes, as they presumably do not identify with the aggressive rules of argumentative engagement and find it difficult to enact a "masculine behavior" of confrontation (Gilbert, 1994). What's more, given expectations that women should behave "politely," their authority is systematically undermined in argumentative situations (Burrow, 2010). For some authors (e.g. Nye, 1990), *any* form of adversariality is stereotypically masculine and intrinsically problematic. For others, the issue is not with adversariality per se, but with the ways in which it happens to be interpreted with excessive, vicious aggressiveness in argumentative situations. Yet others (Govier, 1999) (Aikin, 2011) believe that some amount of adversariality is intrinsic to argumentation, and not necessarily a bad thing.

A more positive appreciation of the role of argumentation in situations of conflict underpins the concept of *agonistic democracy*, which is contrasted with the concept of deliberative democracy. It is motivated by the recognition that some amount of adversariality cannot be entirely eliminated from the political sphere (Wenman, 2013). For agonistic thinkers, arguing in a democracy is not about deliberating towards consensus but about *managing* disagreement, plurality and conflicts of interest. For Chantal Mouffe, for example, "what liberal democratic

politics requires is that the others are not seen as enemies to be destroyed, but as adversaries whose ideas might be fought, even fiercely, but whose right to defend those ideas is not to be questioned.” (Mouffe, 2013, p. 7) The goal is to turn ‘antagonism’ (struggle between enemies) into ‘agonism’ (struggle between adversaries), and argumentation has an important role to play in these practices (Dutilh Novaes, forthcoming).

But what to make of critiques of adversarial models of argumentation? Many of the points made are well taken. Should we therefore abandon argumentation altogether, as an inherently coercive, oppressive mode of interpersonal communication? Or should we aim at developing alternative models which would constrain and regulate excessive displays of aggressiveness? Before proposing reforms to argumentative practices, we should first examine whether current practices do indeed deliver what we expect from them.

## 5. THE OUTCOMES OF ARGUMENTATION

**D**oes argumentation “work”? Does it deliver what we expect from it? When you and your friends engage in an exchange of reasons pro and con the different options, are you eventually able to come to a decision about what to do on Friday that makes everyone (reasonably) happy? In practice, argumentation tends to receive mixed reviews: there is some good news but also quite some bad news. We’ve identified three functions typically attributed to argumentation: to improve beliefs, to promote consensus and social coordination, and to overpower and manipulate (or more optimistically, for agonist thinkers, to manage conflict). Let us now discuss each of them in turn.

Let us start with the putative epistemic function(s) of argumentation: does engaging in argumentation improve one’s beliefs? There is quite some empirical evidence showing that, at least for certain kinds of problems, groups that are allowed to engage in discussion tend to be more successful at finding the right solutions than individual thinkers (Mercier, 2018). The problems where group discussion tends to have a beneficial effect are typically *intellective* problems, that is those that have a demonstrably correct solution within a mathematical, logical, scientific, or verbal conceptual system. For problems for which no generally accepted demonstrably correct answer exists, i.e. evaluative, behavioral, or aesthetic problems—known as *judgmental* problems—the results are not as straightforward.<sup>2</sup>

---

2. “Intellective” and “judgmental” are terms from the group problem-solving literature (Laughlin, 2011).

Consider the following question that was given to participants in an experiment. “Paul is looking at Linda and Linda is looking at Patrick. Paul is married but Patrick is not. Is a person who is married looking at a person who is not married?” Cognitive scientist Hugo Mercier and colleagues studied how reasoners solved this problem, individually and in groups (Trouche, Sander, & Mercier, 2014). Participants in the experiment were given three options to choose from: Yes/No/Cannot be determined. The correct response is “Yes”: if Linda is married, the conclusion follows (Linda is looking at Patrick), and if Linda is not married, the conclusion also follows (Paul is looking at Linda) (considering these as the only two options). Hence, in all possible scenarios, a married person is looking at an unmarried one. But only a small subset of individual reasoners gives this answer (around 20%). Most responses are “cannot be determined,” which is incorrect. By contrast, when solving this problem in groups and debating with peers about which one is the right answer, the rate of correct answers goes to 60%. Thus, discussion and argumentation help find the right answer.

However, it seems that it is only under some special conditions that argumentation will straightforwardly lead to more accurate beliefs. First of all, there is evidence showing that arguments are in fact not a very efficient means to change minds in many real-life situations (Kolbert, 2017) (Gordon-Smith, 2019). People typically do not like to change their minds about firmly entrenched beliefs they hold, and so when confronted with arguments or evidence that contradict these beliefs, they tend either to look away or to discredit the source of the argument as unreliable—a phenomenon known as “confirmation bias” (Nickerson, 1998). In particular, arguments that threaten our core beliefs and our sense of belonging to a group (e.g. political beliefs) typically trigger all kinds of motivated reasoning (Taber & Lodge, 2006) (Kahan, 2017) whereby one outright rejects those arguments without properly engaging with their content.

Relatedly, when choosing among a vast supply of media, people tend to gravitate towards content and sources that confirms their existing opinions, which is one of the factors giving rise to so-called “echo chambers” and “epistemic bubbles” (Nguyen, 2020). Furthermore, some arguments can be deceptively convincing in that they look valid but are not; these are known as *fallacies* in argumentation theory (Hansen, 2019). Because most of us are not very good at spotting fallacious arguments, especially if they are arguments that lend support to the beliefs we already hold, engaging in argumentation may in fact *decrease* the accuracy of our beliefs by convincing us with incorrect arguments supporting false conclusions. In sum, despite the optimism of Mill and many others, engaging in argumentation will not automatically improve our beliefs (even if this may occur in some circumstances) (Dutilh Novaes, 2020a).

What about argumentation as a means to achieve consensus and social coordination? Here too it seems that it is only under specific conditions that



argumentation leads to consensus, as suggested by formal simulations of argumentative situations (Betz, 2013) (Olsson, 2013).<sup>3</sup> In particular, the discussing parties must already have a significant amount of background agreement, especially agreement on what counts as a legitimate argument or convincing evidence (higher-order agreement), for argumentation and deliberation to lead to consensus. Instead, in many real-life situations, argumentation often leads to the opposite result; people disagree with each other even more after engaging in argumentation. This is the well-documented phenomenon of *group polarization*, which occurs when an initial position or tendency of individual members of a group becomes more extreme after group discussion (Sunstein, 2002) (Isenberg, 1986). Indeed, anyone who spends some time on the internet will immediately recognize the dynamics of radicalization triggered by online discussions. Arguably, in part because of the influence of social media, democracies around the world are under threat due to extreme levels of polarization, which (according to some authors at least) make it virtually impossible for functional democratic processes to continue (Talissee, 2019).

Turning now to the third view, is argumentation indeed often used as a means to overpower and manipulate audiences? It depends on what we mean by “argumentation.” Seen as the dispassionate examination of the pros and cons of a particular view, argumentation is not a very effective way to change minds, as mentioned above, and thus presumably not an effective instrument of manipulation. Instead, when two parties disagree and engage in argumentation, more often than not it seems they talk *past* each other rather than *to* each other. Alternatively, they may be trying to “score points” vis-à-vis a presumed audience, which in the day and age of social media translates into “likes”, “shares” and increased popularity.

However, if more broadly construed as including a range of rhetorical devices to enhance persuasion, then argumentation comes closer to the concept of *propaganda*, which is the ultimate instrument to manipulate opinions (Stanley, 2015). To be sure, this does not mean that rhetoric has no place in argumentation that is not propaganda; to the contrary, persuasion, and thus rhetoric, is a fundamental component of argumentation in general. But as already noted by Plato in his critique of the rhetoricians/sophists in the *Gorgias*, persuasiveness can be put to “good” uses (i.e. inform people and let them make up their own minds) as well as “bad” uses, aiming at manipulation and indoctrination. The line between argumentation and propaganda is rather thin indeed.

---

3. There is evidence suggesting that in some organized, moderated contexts, deliberation can in fact lead to satisfactory solutions, for example in the form of compromises (Curato et al., 2017). The key point is that these argumentative environments must be suitably designed and moderated in appropriate institutional settings.

## 6. SOME AREAS OF APPLICATION OF ARGUMENTATION

The previous section painted a somewhat pessimistic picture of what argumentation can do for us. Given these rather bleak prospects, should we give up on argumentation altogether? In this section, I argue that, in specific domains not yet discussed here (political deliberation has already been discussed), argumentation functions reasonably well, and continues to have an important role. By examining the conditions under which this occurs in these specific domains, we may even draw some conclusions on how to improve argumentative processes in other spheres of human life.

Scientific communities arguably offer the best examples of well-functioning argumentative practices. They are disciplined systems of collective epistemic activity, with tacit but widely endorsed norms for argumentative engagement. Indeed, when a scientist presents a new scientific claim, it must be backed by arguments and evidence that her peers are likely to find convincing, as they follow from the application of widely agreed-upon scientific methods. Other scientists will in turn critically examine the evidence and arguments provided, and will voice objections or concerns if they find aspects of the theory to be less convincing (Longino, 1990). Thus seen, science is very much a game of “giving and asking for reasons” (Zamora Bonilla, 2006). Certain features of scientific argumentation seem to ensure its success: scientists see other scientists as *prima facie* peers, and so attribute a fair amount of trust to other scientists by default; science is based on the principle of “organized skepticism” (a term introduced by the pioneer sociologist of science Robert Merton) (Huutoniemi, 2015), which means that asking for further reasons is typically not perceived as a personal attack.

Another area where argumentation is essential is the law, which also corresponds to disciplined systems of collective activity with strict rules for what counts as acceptable arguments and evidence (Feteris, 2017). In litigation (in particular in adversarial justice systems), there are typically two sides disagreeing on what is lawful or just, and the basic decision-making principle is that each side will present its strongest arguments. It is the comparison between the two sets of arguments that should lead to the best decision. While many justice systems around the world are far from perfect and could certainly use reforms,<sup>4</sup> the general principle that legal decisions should be made on the basis of solid argumentation and examination of evidence is arguably indispensable.

Finally, the development of argumentative skills is a fundamental aspect of (formal) education (Muller Mirza & Perret-Clermont, 2009). What is specific

---

4. Season 3 of the podcast ‘Serial’ exposes many of the absurdities of the justice system in the United States.

to argumentation is the focus on *epistemic autonomy*: when presented with arguments, a knower should not simply accept what is being said at face value, but instead should reflect on the reasons offered and come to her own conclusions. Argumentation thus fosters independent, critical thinking, which is an important goal for education. A number of education theorists and developmental psychologists have empirically investigated the effects of emphasizing argumentative skills in educational settings, and the results are encouraging (Kuhn & Crowell, 2011).

## 7. CONCLUSIONS

**V**iews on argumentation, both with respect to its goals and its efficacy, vary widely. This is probably simply a result of the fact that argumentation is in fact many things, indeed a multifaceted phenomenon. This chapter surveyed some influential views on argumentation, both in philosophy and in other relevant fields. We saw that, while argumentation certainly has its limitations, giving up on argumentation completely is too extreme a reaction. Under the right circumstances, an argumentative approach may even facilitate a successful decision-making process on what to do with your friends on a Friday evening.

## BIBLIOGRAPHY

- Aikin, S. (2011). A Defense of war and sport metaphors in argument. *Philosophy and Rhetoric*, 44, 250–272. <https://doi.org/10.5325/phillrhet.44.3.0250>
- Betz, G. (2013). *Debate dynamics: How controversy improves our beliefs*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-4599-5>
- Burrow, S. (2010). Verbal sparring and apologetic points: Politeness in gendered argumentation contexts. *Informal Logic*, 30(3), 235–262. <https://doi.org/10.22329/il.v30i3.3033>
- Cohen, D. (1995). Argument is War...and war is hell: Philosophy, education, and metaphors for argumentation. *Informal Logic*, 17(2), 177–188. <https://doi.org/10.22329/il.v17i2.2406>
- Curato, N., Dryzek, J., Ercan, S., Hendriks, C., & Niemeyer, S. (2017). Twelve key findings in deliberative democracy research. *Daedalus*, 146(3), 28–38. [https://doi.org/10.1162/daed\\_a\\_00444](https://doi.org/10.1162/daed_a_00444)
- Dutilh Novaes, C. (2020a). The role of trust in argumentation. *Informal Logic*, 40(2), 205–236. <https://doi.org/10.22329/il.v40i2.6328>
- Dutilh Novaes, C. (2020b). Reassessing different conceptions of argumentation. In A. Tanesini, & m. Lynch, *polarisation, arrogance, and dogmatism: Philosophical perspectives*. Routledge
- Dutilh Novaes, C. (2021). *The dialogical roots of deduction*. Cambridge University Press. <https://doi.org/10.1017/9781108800792>
- Dutilh Novaes, C. (forthcoming). Who's afraid of adversariality? Conflict and cooperation in argumentation. *Topoi. An International Review of Philosophy*. <https://doi.org/10.1007/s11245-020-09736-9>
- Feteris, E. (2017). Legal argumentation and legal interpretation. In *Argumentation library* (pp. 1–21). Springer Netherlands. [https://doi.org/10.1007/978-94-024-1129-4\\_1](https://doi.org/10.1007/978-94-024-1129-4_1)
- Fishkin, J. (2016). Deliberative democracy. In & M. B. R. Scott (Ed.), *Emerging trends in the social and behavioral sciences* (pp. 1–16). Wiley. <https://doi.org/10.1002/9781118900772.etrds0405>
- Geuss, R. (2019, June 18). *A republic of discussion*. Retrieved from The Point Magazine <https://thepointmag.com/politics/a-republic-of-discussion-habermas-at-ninety/>.
- Gilbert, M. (1994). Feminism, argumentation and coalescence. *Informal Logic*, 16(2), 95–113. <https://doi.org/10.22329/il.v16i2.2444>
- Goldman, A.I. (1994). Argumentation and social epistemology. *Journal of Philosophy*, 91(1), 27–49. <https://doi.org/10.2307/2940949>
- Govier, T. (1999). *The philosophy of argument*. Vale Press.
- Groarke, L. (2016). Informal Logic. In *The Stanford Encyclopedia of Philosophy* (pp. 35–45). Stanford University Press. <https://doi.org/10.4324/9781315253329-9>
- Habermas, J. (1996). *Between facts and norms*. MIT Press.
- Halliday, D., & McCabe, H. (2019). John stuart mill on free speech. In & J. C. D. Coady (Ed.), *The routledge handbook of applied epistemology* (pp. 71–87). Routledge. <https://doi.org/10.4324/9781315679099-6>

- Hansen, H. (2019). Rhetorical ploys and fallacies Rhetorical ploys • Fallacies • Further fallacies. In E. Zalta (Ed.), *Critical thinking* (pp. 125–179). Routledge.  
<https://doi.org/10.4324/9780203482889-10>
- Hundleby, C. (forthcoming). Feminist perspectives on argumentation. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
- Huutoniemi, K. (2015). Peer review: Organized skepticism. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 685–689). Elsevier.  
<https://doi.org/10.1016/b978-0-08-097086-8.03019-1>
- Irani, T. (2017). *Plato on the value of philosophy*. Cambridge University Press.
- Isenberg, D. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50(6), 1141–1151.  
<https://doi.org/10.1037/0022-3514.50.6.1141>
- Kahan, D. (2017). *Misconceptions, misinformation, and the logic of identity-protective cognition*. Elsevier BV. <https://doi.org/10.2139/ssrn.2973067>
- Kidd, I. (2020). Martial metaphors and argumentative virtues and vices. In *Polarisation, arrogance, and dogmatism* (pp. 25–38). Routledge.  
<https://doi.org/10.4324/9780429291395-4>
- Kolbert, E. (2017). *Why facts don't change our minds. The new yorker*.
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, 22(4), 545–552.  
<https://doi.org/10.1177/0956797611402512>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Landemore, H. (2012). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.
- Laughlin, P. (2011). *Group problem solving*. Princeton University Press.
- Lewiński, M., & Mohammed, D. (2016). Argumentation theory. In *The international encyclopedia of communication theory and philosophy* (pp. 1–15). Wiley.  
<https://doi.org/10.1002/9781118766804.wbiect198>
- Longino, H. (1990). *Science as social knowledge*. Princeton University Press.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.
- Mercier, Hugo. (2018). *Reasoning and argumentation* (& V. T. L. Ball, Ed.; pp. 1–6). Wiley.  
<https://doi.org/10.1002/9781118924396.wbiea1513>
- Mill, J. S. (1999). *On liberty*. Broadview Press. (Original work published 1859)
- Mouffe, C. (1999). Deliberative Democracy or Agonistic Pluralism? *Social Research*, 66(3), 745–758.
- Mouffe, C. (2013). *Agonistics*. Verso.
- Moulton, J. (1983). A paradigm of philosophy: The adversary method. In *Synthese library* (pp. 149–164). Kluwer Academic Publishers.  
[https://doi.org/10.1007/0-306-48017-4\\_9](https://doi.org/10.1007/0-306-48017-4_9)
- Muller Mirza, N., & Perret-Clermont, A. (2009). *Argumentation and Education*. Springer.
- Nehamas, A. (1990). Eristic, Antilogic, Sophistic, Dialectic: Plato's Demarcation of Philosophy from Sophistry. *History of Philosophy Quarterly*, 7, 3–16.  
<https://doi.org/10.1515/9783110365870.67>

- Nguyen, C. (2020). Echo Chambers and Epistemic Bubbles. *Problem of Reductionism in Science*, 17(2), 141–161. <https://doi.org/10.1017/epi.2018.32>
- Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Norman, A. (2016). Why we reason: Intention-alignment and the genesis of human rationality. *Biology & Philosophy*, 31(5), 685–704. <https://doi.org/10.1007/s10539-016-9532-4>
- Nozick, R. (1981). *Philosophical explanations*. Routledge.
- Nye, A. (1990). *Words of Power*. Routledge.
- Olson, K. (2014). Deliberative democracy. In B. Fultner (Ed.), *Jürgen Habermas—Key Concepts* (pp. 140–155). Acumen Publishing Limited. <https://doi.org/10.1017/upo9781844654741.008>
- Olsson, E. (2013). A Bayesian simulation model of group deliberation and polarization. In F. Zenker (Ed.), *Bayesian argumentation* (pp. 113–133). Springer Netherlands. [https://doi.org/10.1007/978-94-007-5357-0\\_6](https://doi.org/10.1007/978-94-007-5357-0_6)
- Peter, F. (2016). The Epistemology of Deliberative Democracy. In K. Lippert-Rasmussen, K. Brownlee, & D. Coady (Eds.), *A Companion to Applied Philosophy* (pp. 76–88). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118869109.ch6>
- Rooney, P. (2012). When philosophical argumentation impedes social and political progress. *Journal of Social Philosophy*, 43(3), 317–333. <https://doi.org/10.1111/j.1467-9833.2012.01568.x>
- Stanley, J. (2015). *How propaganda works*. Princeton University Press.
- Sunstein, C. (2002). The law of group polarization. *Journal of Political Philosophy*, 10(2), 175–195. <https://doi.org/10.1111/1467-9760.00148>
- Taber, C., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Talisso, R. (2019). *Overdoing democracy*. Oxford University Press.
- Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.
- Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, 143(5), 1958–1971. <https://doi.org/10.1037/a0037099>
- Van Eemeren, F., & Grootendorst, R. (1992). *Argumentation, communication, and fallacies*. Routledge. <https://doi.org/10.4324/9781315538662>
- Van Eemeren, F., Grootendorst, R., Johnson, R., Plantin, C., & Willard, C. (1996). *Fundamentals of argumentation theory*. Routledge. <https://doi.org/10.4324/9780203811306>
- Walton, D., & Krabbe, E. C. W. (1995). *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press.
- Wenman, M. (2013). *Agonistic democracy*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511777158>
- Zamora Bonilla, J. (2006). Science as a persuasion game: An inferentialist approach. *Problem of Reductionism in Science*, 2(3), 189–201. <https://doi.org/10.3366/epi.2005.2.3.189>

JOHN STUART MILL



## PLATO

### GORGIAS

---

**T**he *Gorgias* is one of Plato's early dialogues (circa 380 BCE), one of the many where Socrates features as the main character. Socrates engages in conversation with three interlocutors who identify as sophists: Gorgias, one of the most respected sophists/rhetoricians of the time, and two of his disciples, Polus and Callicles. The initial topic under discussion is the very definition of rhetoric. Socrates argues that, although it can be put to good use, in practice rhetoric is no more than flattery: telling people what they want to hear, rather than aiming at the truth, in order to manipulate their opinions and views. Rhetoric thus only serves to convince the ignorant, whereas philosophy promotes a virtuous life. The conversation then turns into a discussion of what it means to lead a good life. Callicles, the most extreme and outspoken of the three, maintains that a good life is merely a life of maximal pleasure, where the stronger rightfully dominates the weaker. Despite his best efforts, Socrates is unable to change Callicles' mind with his arguments. The dialogue as a whole offers deep reflections on how to engage in fruitful argumentation, and the limits and scope of rational discourse.

---



---

# THINKING ABOUT REALITY

---

René van Woudenberg

**This is the Question:  
To Be or Not to Be?**

# THIS IS THE QUESTION: TO BE OR NOT TO BE?

## *An Introduction to Ontology*

René van Woudenberg  
[r.van.woudenberg@vu.nl](mailto:r.van.woudenberg@vu.nl)

### 1. INTRODUCTION

**O**ntology designates the area of philosophical reflection that circles around the words that occur twice in the one Shakespeare quotation that everybody knows even when they haven't read *Hamlet*: 'to be', the equivalent of which is 'to exist', and is usually considered to be a part of metaphysics. The central ontological question, as the American philosopher Willard V.O. Quine has famously said, is "What is there?" Part of an answer to which could be "there is the number 5" or more generally, "there are numbers" (or "numbers exist"). Part of an answer could also be "there are rocks" or more generally, "there are material objects (they exist)."<sup>1</sup>

If Quine has formulated the central ontological question, then there is a question that logically precedes it, i.e. "What *is* existence?" or, alternatively, "What *is* it for something to exist?". Since the question is about the meaning of ontology's central concept, it can be called the "meta-ontological question". It is natural to think that we need an answer to the meta-ontological question if we are to fruitfully approach the ontological question. Accordingly, in this chapter I will consider both the meta-ontological and the ontological question. Within the

---

1. The ontological question arises across philosophical disciplines: in the philosophy of language ("Do meanings exist?"), in moral philosophy ("Do obligations exist?"), in the philosophy of science ("Do unobservable entities exist?"), in the philosophy of mind ("Do beliefs exist?"), etc.

space of this chapter, the treatment will necessarily be brief, but it will be enough to get a sense of the main topics in ontology, as well as of various approaches to the meta-ontological question.

## 2. IS THERE A PROPERTY THAT EVERYTHING THAT EXISTS HAS?

I start with the meta-ontological question “What is it for something X to exist?” or, alternatively, “Is there a feature or property such that everything that exists has *it*, and that in order for something to exist it has to have *it*?” Let us consider the following answer: a thing exists provided it is physical—and let us call this view *physicalism*. “To be” and “to be physical”, according to friends of physicalism, are two names for the same property. If we say that “X exists”, then, the physicalist holds, we say no more and no less than that “X is a physical thing”.

This answer is instructive because of two different responses it can (and should) elicit, a semantic and an epistemological one. The semantic response - having to do with the meaning of words - is as follows: the two terms “to be” and “to be physical” might refer to the same set of objects, yet can have different meanings. We must distinguish between the meaning of a term and the objects to which the term applies. The meaning of the word ‘dog’ lists properties that an animal must have in order to be a dog. This is called the intension of the word ‘dog’. This is different from its extension: that is the set of objects to which the term applies, the set of dogs. Two terms can have the same extension but different intensions, e.g. the terms ‘having exactly and only three sides’ and ‘having exactly and only three angles’ have obviously different intensions: they are intensionally diverse. But they are extensionally equivalent, i.e. the set of things to which the one term truly applies is identical to the set of things to which the other term truly applies. One response to physicalism is along these lines: it says that even if “to be” and “to be physical” are extensionally equivalent, this doesn’t mean that these terms are also intensionally equivalent. If the two were intensionally equivalent, they would be synonyms in the way *taxi* and *cab* are synonyms. But “to be” and “to be physical” just aren’t synonyms. When we say “X is a taxi” and “X is a cab” we say the same thing about X. But when we say “X is” and “X is physical”, in the latter statement something is predicated of X that just isn’t predicated of X in the former statement. “To be” and “to be physical”, unlike “is a taxi” and “is a cab”, are not intensionally identical, they aren’t synonyms. Indeed, Hamlet didn’t say that the question is to be physical or to not be physical!

But the physicalist may still be right: “to be” and “to be physical” may be extensionally equivalent—that is, everything that exists may *in fact* be something physical. This leads up to the second, *epistemological*, response that physicalism should elicit, viz. “how did you, physicalists, figure out that everything that exists is something physical?—how did you arrive at the conclusion that ‘to be’ and ‘to

be physical' are extensionally equivalent?" This is by no means an easy question for physicalists to answer. For they surely have *not* figured this out in the way chemists *have* figured out that water is H<sub>2</sub>O, and astronomers have figured out that the Eveningstar is the Morningstar—that is, not through some form of scientific research.

Physicalism isn't the only view that elicits (and *should* elicit) the two responses mentioned. *Every* view that says that "to be" just is "to be F" (where "to be F" stands dummy for possible views like "to be in time-space", "to stand in causal relations to other things", "to be concrete", "to be mental") should elicit the semantic and the epistemological responses.

The conclusion we can draw is that any claim to the effect that "to be" is "to be F" is problematic. That is to say, any answer to the question that is the title of this section ("Is there a property that everything that exists has?") and that has the form "yes, there is such a property, viz. property F", is problematic. What can, perhaps, be said is that everything that exists has the property of existence—and that this is the best answer we can give to the question at hand. I say "perhaps", for the 18<sup>th</sup> century philosopher Immanuel Kant has famously argued that this cannot be right, since existence, as he contends, is not a property.

### 3. IS "EXISTENCE" OR "BEING" A PROPERTY?

**K**ant's denial that 'existence' is a property can be illustrated as follows. Suppose we are asked to sum up the properties of the Eiffel tower. Then we will likely mention that it is made of mainly iron, that it is over 300 meters high, that it has four legs, that it is located on the left bank of the Seine, that its official opening was in 1889, and that in recent times it is yearly visited by over 6 million people. But shouldn't we add to this list "that it exists" as well: isn't "existence" one of the properties of the Eiffel tower too? As indicated, Kant says we should not. His reason is this: in order for something (whether it is the city of Königsbergen, the President of the U.S., or the Eiffel tower doesn't matter) to have properties *at all*, it should exist. Only things that exist can have properties. And hence, Kant held, "existence" just isn't a property.

It would seem that Kant's line of thought is unobjectionable, except for the last step. It is unobjectionable that "existence" cannot be added to the list of properties of the Eiffel tower. It is unobjectionable that the Eiffel tower can only have properties because it exists. But it *is* objectionable to conclude from this that "existence" is therefore not a property.

Let me explain. Kant held that "existence" is not a property of *things*—i.e. of *substances* like the Eiffel tower. But from this it cannot be concluded that it is

not a property *at all*. For we can think of “existence” as a property not of *things*, but as a property of a set of *properties* (and this may even be what Kant himself held). For example, consider the following set of properties: “is made of mainly iron”, “is over 300 meters high”, “has four legs”, “is located on the left bank of the Seine”, “was officially opened in 1889”, and “is in recent times yearly visited by over 6 million people”. Then we can say that to affirm that the Eiffel tower exists, is to affirm that this set of properties has the following property: *the properties in this set are the properties of one thing*. This illustrates how we can think of “existence” as a property—a property not of individual things, but of sets of properties.

We should take note of the fact that this explanation of how we can think about “existence” as a property, included a numeral, i.e. the number “one”. The 19<sup>th</sup> century German mathematician and philosopher Gottlob Frege made the striking and valuable observation that there is a deep connection between the notion of “existence” and the notion of “number”. His observation has been captured in the slogan that *affirmation of existence is denial of the number zero*. What the slogan aims to express can best be explained by examples. To say that planets exist, is to say that the number of planets is not zero. To say that there are female prime ministers, is to say that the number of female prime ministers is not zero. And if we say that there are no winged horses, we say that the number of winged horses is zero.

Frege’s slogan gives us a handle on how to understand affirmations of the existence not only of *kinds* of things (planets, prime ministers), but also of individual things, such as the Eiffel tower or the painter Holman Hunt. The slogan suggests the following. To say that the Eiffel tower exists, is to say that the number of things identical with the Eiffel tower is 1. And to say that Holman Hunt exists, is to say that the number of things identical with Holman Hunt is 1.

Now, if we take Frege’s slogan on board, we can even think about existence as a property of individual things. If the Eiffel tower exists, then the Eiffel tower’s existence is this property: *the number of things identical to it is 1*. And if there is a tallest man in South Dakota, then that man’s existence too is the following property: *the number of things identical to it is 1*.

Seen from this perspective, the conclusion, then, is that Kant’s conclusion that “existence” is not a property must be qualified.

#### 4. ARE “BEING” AND “EXISTENCE” THE SAME NOTION?

In the opening paragraph I said that “to be” and “to exist” are synonyms, have the same meaning. But this has been denied by, among others, the late 19<sup>th</sup> century Austrian philosopher Alexius Meinong. He claimed that “to be” and “to exist” are different notions. If he is correct, this would mean that the sentences “There are black holes” and “Black holes exist” say different things. *The distinction enabled him to say that there are things that don’t exist*—examples of which would be the President of the UK and the 24<sup>th</sup> child of Henry Kissinger. The President of the UK should then thought to “be”, although his or her “existence” should be denied (the UK has no President, and never has had one). Likewise, the 24<sup>th</sup> child of Kissinger should be thought to “be”, albeit that his or her “existence” should be denied (Kissinger having no 24<sup>th</sup> child).

One argument philosophers have advanced to buttress the distinction between “being” and “existence” is that it enables us to say such things as “The President of the UK doesn’t exist”—which, it is claimed, can only make sense if the President of the UK can be referred to, which requires in turn that the President of the UK *is*. The thought is that we can only refer to things that *are*, even if the things we refer to do not exist.

But this line of thought can be resisted, and even in a fairly convincing way, as the famous 20<sup>th</sup> century philosopher and Nobel laureate Bertrand Russell has showed. The sentence “The President of the UK doesn’t exist” should not be thought to refer to anything at all—and certainly not to a non-existing head of a non-existing republic. Why not? Because that sentence can be paraphrased into an equivalent sentence (one with exactly the same meaning) that does not even *seem* to carry the entailment that there is something that doesn’t exist. The paraphrase is this: “Nothing is the President of the UK”. Or, formulated in the way suggested by Frege: “The number of things that is the President of the UK is zero”. These sentences are equivalent to the sentence we started with, but do not suggest that the President of the UK, somehow, *is*.

Another argument that has led some philosophers to believe there is a distinction between being and existence (and hence that there are things that do not exist) has to do with intentionality. Intentionality, Franz Brentano famously said, is “the mark of the mental”. By this he meant that when we think, we think *about something*—our thinking is, when we think, directed to an object. When we think about the Swiss mountains, or think that  $5+7=12$ , our thoughts are directed to the Swiss mountains and a particular sum respectively. But now note that it is possible to think about things that do not exist. You can think about the sister you never had, or about Scrooge and Marley –two character in Charles Dickens’

*A Christmas Carol in Prose*—none of who exist. But they *are* there for you to think about nonetheless.

This line of thought too can be resisted. One could, for example, think that it is just wrong to say that each and every thought has an object. *Some* thoughts have objects, for example your thoughts about the Swiss mountains, and your thoughts about the Eiffel tower. But not *all* thoughts have objects—for instance your thoughts about the sister you never had, and about Scrooge and Marley. They are as little directed at something as an archer’s arrow is directed at something when the archer, mistakenly, takes a cloud formation for a dove.

There is, however, something unsatisfactory about this response. For it seems that your thoughts are really *about* something, and that they *do have objects* when you think about the sister you never had, and about Scrooge and Marley. To be sure, the sister you never had is not a person of flesh and blood, someone who breathes and eats, someone who has thoughts and feelings. There is no person of flesh and blood who is the sister you never had. The number of persons that are the sister you never had is zero. But nonetheless—there is the *fantasy image* of the sister you never had. And *that fantasy image* does exist and *it* is the object of your thoughts when you think about the sister you never had. The same can be said about Scrooge and Marley: there are no persons of flesh and blood that are Scrooge and Marley, but there are *fictional characters* that are Scrooge and Marley. And fictional characters exist, just as the person of flesh and blood who is your neighbor exists. When you think about Scrooge and Marley, the object of your thoughts is not persons of flesh and blood, but fictional characters.

If this line of reasoning is correct, then we should conclude that the arguments for distinguishing between being and existence fail, and that as yet we have seen no reason for making this distinction. (And we should also note that in ordinary language we just use “being” and “existing” as synonyms). But this line of reasoning does give rise to a new question:

## 5. HAS “EXIST” MULTIPLE SENSES, OR ONE?

**T**his question arises, for you may think (and I have heard it said many times) that in the following sentences the word “exist” and its synonyms and derivations have different senses:

1. Stones exist
2. Cells exist
3. There are cows in the meadow
4. There are hotels in Paris
5. There is a prime number between 4 and 6



6. God exists
7. There is a principle that says that if two things are equal to a third thing, they are equal to each other
8. There is such a thing as the virtue of fidelity

You may think that when we say about stones that *they* exist, we say something different about them than when we say about cells that *they* exist. Likewise, you may think that for cows to exist is something different than what it is for hotels in Paris to exist. And again, you may think that what it is for a prime number or a principle to exist is very different than what it is for a virtue or for God to exist. Your question is, thus, whether in each of these statements the word “exist” is used in the same sense. Or, to put the same question in yet another way, your question is whether “exists” in all these sentences is used *univocally* (i.e. in the same sense), or *analogically* (i.e. in somewhat related senses) or *equivocally* (i.e. in unrelated senses).

One possible answer, one that appeals mostly to philosophers in the continental-existentialist tradition (see Corijn van Mazijk’s chapter on “Examining Life”), has it that in the eight sentences on the list “exist” is used in different, though perhaps related senses—but surely not univocally, not in the same sense. These philosophers are attracted to the notion of “ways of being” and they like to say that stones and cells, cows and hotels, numbers, principles and God all have their own, and possibly unique, “way of being”, that they “exist” in different ways, and hence that “exist”, when said of the items on the list, is used in analogical senses. It is incumbent upon philosophers who think this way to explain what these analogically related senses of “exists” *are*. And when they seriously try to do so, they will have to say such things as: “well, stones are lifeless but cells are alive—hence stones exist in another way, have another ‘mode of being’, than cells. And cows are concrete, while numbers are abstract—hence, cows exist in another way that numbers, they have different ‘modes of being’. And so on for the other items on the list.”

There is reason to think, however, that this line of thought is not the best one to have. For there is an elegant alternative: all the items on the list, assuming (if only for the sake of argument) that they are real, are items with radically different *natures*. Let us say that two things are of a different nature if the set of essential properties that one of the things has is not identical to the set of essential properties that the other thing has. And a property is an essential property of a thing, provided the following is the case: if the thing would lose that property, it would stop being that thing. For example: being odd is an essential property of the number 9—it is a property that the number 9 could not lose without stopping to be that number. Also, being a person is an essential property of you reader—it is a property that you have, and continue to have through all the changes you

have gone through. Loss of that property would mean the termination of your existence. Now, a stone has a nature that differs from that of a cell, because the set of essential properties of a cell includes the property of being capable of replication—which is not an element in the set of essential properties of a stone. And so it goes for all the items on the list: I selected them because each has a nature that differs from the nature of all the other items on the list. So, the list refers to items with radically different natures. Now we *know* that things of different natures exist: we know, for example, that stones exist and we also know that cells exist etc. But if all of these things, with their different natures, exist, they exist in the one and only sense that the word “exist” has. So, on this view, the word “exist” (and its synonyms and derivatives) is used univocally throughout the list.

Why should we prefer the latter view? I offer two reasons. First, whereas the first makes use of one unclear notion of “mode or way of being”, the second uses two intuitively very clear notions, viz. (an item’s) “nature”, and (that item’s) “existence”. And it is much better to work with two clear notions than with one unclear one. Second, it is often said that there are close structural parallels between (what we can say about) the notion of “existence” and the notion of “truth”. There is *one* notion of truth—and the traditional way to unpack it is by saying that a statement is true, if and only if what the statement says to be the case, actually is the case. Nothing more is needed and nothing less will suffice. The statements “stones do not reproduce themselves” and “cells reproduce themselves” are both true, even though the first statement states a physical fact and the second a biological fact—and physical facts and biological facts have very different natures. Yet, both statements are true in the one sense that “true” has. This is analogous to saying that stones and cells exist in the one sense that “exist” has, even though stones and cells have radically different natures.

## 6. THE EXISTENCE OF UNIVERSALS

So far this chapter dealt with “meta-ontological” matters. We now consider two “ground-level” ontological questions. We can enter the first by reflecting on the following fact: there are yellow flowers, yellow clothes, and chicken egg’s yolks are yellow too. We can state this fact by saying that there is a property, viz. being yellow, that is a property of many flowers, of many clothes, and of all chicken egg yolks. Also, Oslo is north of Amsterdam, Belo Horizonte is north of Rio de Janeiro, and Vladivostok is north of Pusan. We can state this fact by saying that there is a relation, viz. the “is north of”-relation, in which these ordered pairs of cities stand. Similarly, your neighbour is a human being, your aunt Elizabeth is a human being, and the King of Belgium is a human being. We

can state this fact by saying that there is a kind, viz. the kind “human being” to which all three persons belong.

These statements may sound unremarkable, but there is something truly amazing about them. For they say that there are properties, that there are relations, and that there are kinds. And this, as was suggested before, is equivalent to saying that properties, relations, and kinds *exist*. What is amazing about this is that while many people, when asked to give an inventory of what exists, will say “Well, there are inanimate things like rocks; there are living things like gazelles; there are soccer clubs like Tottenham Hotspur; there are banks like the Royal Bank of Scotland”, but few will add “and there are properties, relations, and kinds”. What is also amazing is that, if properties, relations and kinds truly exist, numerically *different things* can have *the very same* property—which means that one the same thing (property, relation, kind) can be present at different locations, and at different times.

The things that have the properties, and stand in relations or that belong to kinds are usually called *particulars*. This contrasts with the properties themselves. Relations and kinds that are called *universals* as they are shared by numerically distinct particulars. Since properties, relations, and kinds exist, all of which are universals, it follows that *universals* exist. Or so the so-called ‘metaphysical realist’ holds. Universals, unlike particulars, are ‘repeatable’ entities. In the technical terminology that is often used, universals are ‘instantiated in’ particulars, and also: particulars ‘exemplify’ universals (and particulars ‘belong’ to kinds).

Universals are, moreover, deemed ‘abstract’, by which the realist means that they have no location in spacetime. Universals are often also held to exist even when they are not instantiated in any particular (or even if there is no particular that “belongs” to a kind). An example that may illustrate this idea is the property-universal of *being a Dutch female prime-minister*. This property, so far, is uninstantiated. But it exists. Similarly for relations and kinds.

Nominalists disagree with all of this. They hold that only concrete particulars exist and deny the existence of universals. So they hold that the yellow of this flower is different from the yellow of the egg yolk. They likewise hold that the “being north of” relation between Oslo and Amsterdam is different from the one between Belo Horizonte and Rio de Janeiro. Again, they hold that your neighbour and your aunt Elizabeth don’t really belong to the same kind (i.e. the kind “human being”) as there are no kinds. Rather, nominalists say, the yellow of the flower and the yellow of the yoke *resemble* each other, the spatial relation between Oslo and Amsterdam *resembles* that between Belo Horizonte and Rio de Janeiro, and your neighbour and your aunt *resemble* each other in various respects.

This way of stating the nominalist view is not without problems, as it involves a universal itself—the universal of “resembling”. But nominalists have responses to this that are worth while studying.

## 7. WHY IS THERE ANYTHING AT ALL (AND NOT NOTHING)?

Let us now move to perhaps the most baffling of all ontological questions. We know that cells and horses exist, and also that mountains, persons, societies, the Earth, the solar system, and the universe exist. And once we start thinking about the fact that all these things exist, it may strike us in a forceful way. Isn't it incredibly strange, even mysterious, that there are things *at all*? Wasn't it possible that the number of things was zero, that there was nothing, really nothing, not even a vast empty space? The philosopher Leibniz therefore urged us to address the question “Why is there something, and not nothing?”<sup>2</sup> which is a request for an explanation of the fact that there exists anything at all (so his question is not why horses exist, or mountains, or the Earth, etc.).

There is something remarkable about this question that we must note before we can consider possible answers. Normally when we explain the existence of a particular kind of things, we refer to things that are *not* of that kind. For example, when we want to explain why there is life on Earth, we refer to the situation in which there was Earth, but no life on it yet, and we point to factors in that situation that led to the emergence of life on Earth. More generally, we normally heed the following Explanatory Principle: for the explanation of the existence of Fs we refer to things that are *not* Fs but that can held to be causally responsible for Fs coming to be.

If we return to Leibniz's question, and try to apply the Explanatory Principle, we seem to get stuck. For if we want to explain why there is anything at all (and not nothing), the principle dictates that we refer to what is *not* anything, i.e. to what is *not* something—which means that it dictates that it refers to *nothing*. But nothing is nothing. Nothing is not something—hence not something with properties to which we can refer when we try to explain why there is anything at all. This goes to show that there is something rather strange about Leibniz' question: it cannot be answered in a way that conforms to the Explanatory Principle.

For this reason some have thought there is something wrong with the question, and hence that we should not even attempt to answer it. The wrongness might be thought to exist in this: ‘explanation’ is a term that makes sense in very specific contexts, e.g. the contexts in which we ask what explains *this or that particular phenomenon*. But outside such very specific contexts, asking for explanations

---

2. See his “On the ultimate origination of things” (1697) in Leibniz 1971: 337-351.

makes no sense. Still, to many the question “Why is there anything at all?” makes sense, and they have to work towards an answer without wielding the Explanatory Principle. Let us consider a few attempts.

One answer that may come to mind is that God’s existence explains why there is anything at all, God being the creator of all that is. But this answer won’t do. For Leibniz’s question is why there is anything at *all* and thus demands an explanation of God’s existence (if God exists) as well. This answer leaves God’s existence (if God exists) unexplained.

Another answer that has come to many minds is that the fact that there is anything at all is due to chance, or more sophisticated: it is due to wholly indeterministic ‘quantum fluctuations’. But this won’t do either. For chance is a property of events, and hence there must be events if there is to be chance. But the existence of events is thus left unexplained. Likewise for quantum fluctuations: if quantum fluctuations have led to the existence of anything at all, they must have been there, and hence their existence is left unexplained.

Let us consider three possible answers that don’t seem to suffer from the problem that they leave something unexplained. The first one traces back to St. Anselm, and involves the notion of *necessary existence*. According to this line of thinking, some things that exist *could have not existed*—their existence is not necessary but contingent. Examples of contingently existing things include the Eiffel Tower, you and me, and the Kingdom of Sweden. The Eiffel Tower doesn’t exist necessarily—it could never have been built, and so, *mutatis mutandis* for the other items. Contingently existing things contrast with necessarily existing things. Something O exists necessarily provided it is not possible for O not to exist. If O exists necessarily, then nothing whatsoever can, so to say, push O out of existence.

Now think back to Leibniz’s question, why is there anything at all (and not nothing), and keep in mind the notion of necessary existence. Then we can see that the following is at least *an intelligible answer* to the question: there are things (and there are not no things) because there is something that exists necessarily and this necessarily existing thing, somehow, explains the existence of contingent things. So, why is there something and not nothing? Because there is something that exists necessarily. These thoughts form the backbone of the so-called ‘modal ontological argument’ (‘modal’ because the argument contains the modal notions of possibility and necessity) inaugurated by the philosopher Norman Malcolm.

In the Western philosophical tradition this necessarily existing thing was often identified with the Christian God. But this requires thorough argumentation—and it is not necessary to make the identification in order to present the conclusion

of the modal ontological argument (“There is something that exists necessarily”) as at least an intelligible answer to Leibniz’s question.

A second answer to Leibniz’s question that may not suffer from the “unexplained rest” problem goes by the name of the “axiarchic view”. The train of thought here is that of all the countless ways that the whole of reality could have been, one is the best. And it next suggests that reality as it is, is that best of all possible worlds (to use another famous diction of Leibniz: the actual world is the best of possible worlds). The final part of the train consists in suggesting that it is *no accident* that the actual world is the best possible world. The actual world, so the suggestion goes, is actual *because* it is the best.

Friends of the axiarchic view (John Leslie, for example), inevitably, face problems. One problem for them is that the view offers an explanation that is so utterly different in kind from other explanations that we accept and love—like the explanation of why iron expands when heated. But then again, *any* answer to Leibniz’ question will contain elements that are utterly different from what we find in other explanations that we accept and love.

This holds true for a third response to Leibniz question as well. This response says that the fact that there exists anything at all, is a *brute* fact, so a fact that defies explanation of any sort. Brute facts are facts for the obtaining of which there is no reason; there is no reason why the fact obtains as little as there would be a reason if the fact would not obtain. This is a response, not an answer to Leibniz’ question. The response is an acknowledgement of the fact that the human intellect is running out of its depths here, and that we must, perhaps grudgingly, accept that existence is an impenetrable mystery.

Again, many, notably friends of scientism, do not like this response. They hold, very roughly, that if science cannot answer a question, there must be something wrong with it. But then again, scientism itself is of questionable reputation.

The American philosopher Robert Nozick once said that Leibniz’ “question cuts so deep ... that any approach that stands a chance of yielding an answer will look extremely weird. Someone who proposes a non-strange answer shows he didn’t understand the question. Since the question is not to be rejected, though, we must be prepared to accept strangeness or apparent craziness in a theory that answers it.”<sup>3</sup> I think this is well put.

---

3. Nozick 1981: 116.

## 8. ENVOI

In this chapter, then, I have discussed a number of (as we can now see: interrelated) meta-ontological questions: Is there a property that everything that exists has? Is existence a property? Are being and existence synonyms? Are there things that don't exist? Next we discussed, however quickly, two ontological questions: Do properties, relations, and kinds exist? and Why is there anything at all? These are some of the main questions that are being discussed within the branch of metaphysics called ontology.

Can we see or expect new developments—either new questions, or new approaches? With all the provisos one should make here, I think we should expect ontological discussion to turn to the following questions: “What is the ontological status of the internet and of virtual reality”? (see the chapter on the philosophy of technology) “What is the ontological status of persons?” “What is the ontological status of race, of gender as well as of ‘nation’?” (see the chapter on political philosophy), and “What is the ontological status of values?” (see the chapter on meta-ethics.)

By questions about the “ontological status” of something X, I mean questions about whether or not X exists, of what metaphysical nature X is (if any), whether or not X can be “reduced” to other items that are supposed to exist, and whether or not X is a mere illusion. As to new approaches I see and expect a further development of views according to which the ontological question (so the question: what exists?) should be answered by reference to natural science alone, with a special place reserved for evolutionary theory. I also expect a further development of what are called “scientific” ontologies. I expect moreover the further development of explicitly feminist approaches to ontological questions. Finally, I expect the continued discussion of Leibniz’s question by science-inspired means and concepts, such as the multiverse theories and structuralism.

## BIBLIOGRAPHY

The works mentioned below are all truly excellent and often amazing. Armstrong, Aune, and Lowe target undergraduate students, Loux and Crisp students at Master and PhD levels. The volumes by Lewis, Van Inwagen and Chalmers (a.o.) contain rather technical papers and are, accordingly, less readily accessible—but they are of the highest quality. A collection of historical readings on the ontological argument is Plantinga 1965. A subtle discussion of the axiarchic view is Parfit’s “Why Anything? Why This?” in Van Inwagen & Zimmerman 2008, that also contains an interesting reply to the view by Richard Swinburne. A discussion of “scientism” (which should not be thought of as a pejorative term—it is adopted by its proponents as a badge of honor) is the edited volume by De Ridder et al 2018.<sup>4</sup>

Armstrong, D. M. (1989). *Universals. An opiated introduction*. Westview Press.

Aune, B. (1985). *Metaphysics. The Elements*. University of Minneapolis Press.

Chalmers, D. J., Manley, D., & Wasserman, R. (Eds.). (2009). *Metametaphysics: New essays on the foundations of ontology*. Clarendon Press ; Oxford University Press.

de Ridder, J., Peels, R., & van Woudenberg, R. (Eds.). (2018). *Scientism: Prospects and problems* (Vol. 1). Oxford University Press.

<https://doi.org/10.1093/oso/9780190462758.001.0001>

Leibniz, G. W. F. (1971). *The monadology and other philosophical writings* (R. Latta, Trans.). Clarendon Press., <https://doi.org/10.5962/bhl.title.27631>

Lewis, D. (1999). *Papers in metaphysics and epistemology*. Cambridge University Press.

Loux, M.J., & Thomas, M. (2017). *Metaphysics. A Contemporary Introduction* (4th ed.). Routledge. [10.4324/9781315637242](https://doi.org/10.4324/9781315637242)

Lowe, E. J. (2002). *A survey of metaphysics*. Oxford University Press.

Nozick, R. (1981). *Philosophical explanations*. Routledge.

Plantinga, A. (forthcoming). *The ontological argument, from St. Anselm to contemporary philosophers*.

van Inwagen, P., & Zimmerman, D. W. (Eds.). (2008). *Metaphysics: The big questions* (2nd, rev.expanded ed.). Blackwell Publishing

van Inwagen, P. (2014). *Existence. Essays in Ontology*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107111004>

---

4. For comments on an earlier version of this paper many thanks to Valentin Arts, Lieven Decock, Tamarinde Haven, Martin van Hees, Lodi Nauta, Rik Peels, and Jeroen de Ridder.



W. V. O. QUINE



## WILLARD VAN ORMAN QUINE *ON WHAT THERE IS*

---

One of the most influential papers in ontology is W. V. O. Quine's "On What There Is?", published in the *Review of Metaphysics* 1948/49. The paper has a critical and a constructive part. Quine criticizes a Platonist ontology for being far too rich: it contains entities that he finds it hard to accept as real, notably universals (properties like "being yellow"; relations like "being north of" and kinds like "being human") and nonexistent objects like Pegasus, the winged horse of Greek mythology. This critical approach reflects Quine's predilection for "desert landscape ontologies" that he confessed to elsewhere. In the constructive part Quine develops a method to decide which entities we should believe exist—that is he develops a method for finding one's "ontological commitments". The method is to consider all the things you believe to be true, and then see what statements of ontological commitments follow, by the rules of logic, from these beliefs. Quine urges that many statements that seem to commit us to the existence of, say, Xs, don't really do so, as those statements can be "paraphrased" into statements that have the same content but do not commit us to the existence of Xs. For example: the statement that there is a hole in the donut, seems to commit us to the existence of holes, which is puzzling, as holes are nothings. The statement, however, can be paraphrased in a way that has no such commitment: The donut is perforated.

---

Eddo Evink

**The Highest Principles?  
The Death and Rise of  
'First Philosophy'**

# THE HIGHEST PRINCIPLES? THE DEATH AND RISE OF 'FIRST PHILOSOPHY'

## *An Introduction to Metaphysics*

Eddo Evink  
[c.e.evink@rug.nl](mailto:c.e.evink@rug.nl)

### 1. INTRODUCTION

**P**hilosophers never stop asking questions. These questions often concern very fundamental issues: How can we make a difference between right and wrong? What can we know about the world we live in? What is knowledge? What is truth?

As this volume demonstrates, philosophy has many subdisciplines each of which concerns itself with some of these questions. Ethics addresses issues of right and wrong, epistemology concerns the nature of knowledge, anthropology is about the nature of man, etc. Is philosophy therefore an aggregate or even a collection of different disciplines rather than a unified whole? Some philosophers contend that it is in fact a coherent whole because the various branches of philosophy all draw upon and are rooted in the most basic and fundamental of all philosophical disciplines, namely metaphysics. After all, it is metaphysics that inquires after the principles of reality as such: What is a being? Is reality rationally or systematically ordered? What would be the founding principle of this or any order?

The Greek philosopher Aristotle calls this foundation 'first philosophy'. When his lectures were collected, the lectures on 'first philosophy' were grouped after the lectures on physics and therefore they were called *metaphysics* [*ta meta ta*

*physika*], which literally means ‘the [work] after physics’. ‘Metaphysics’ can also be translated as ‘beyond physics’, referring to the most general characteristics of reality (e.g., what is a being?) that transcend the qualities of natural things as they are researched in physics (e.g. what is the speed of sound?). In Aristotle’s view, asking such fundamental questions is essential. Moreover, people have a natural capacity to inquire knowledge and a desire to know, as Aristotle signifies in the opening of his text: “All men by nature desire to know” (Aristotle 1984, II 1552).

Through the ages, metaphysics was indeed seen as the ‘first philosophy’, as the basis and foundation of all other thoughts and reflections. From the nineteenth century onwards, however, it has been regarded more and more as a branch of philosophy, the reliability of which needed to be critically questioned (Kant 1999) or as a branch of philosophy with unanswerable questions that would be better left behind (Nietzsche). Some philosophers have even straightforwardly called metaphysics impossible or useless (Dilthey 1989; Carnap 2003). Philosophers of the so-called Vienna Circle approached it with downright hostility. But the questions about the world as a whole and about life as such return again and again. Metaphysics cannot be dismissed so easily. Is there still a place for metaphysics in contemporary philosophy? And if so, what would it look like?

To address these questions, I shall first sketch some prominent features of the metaphysical tradition as it was developed in ancient Greek philosophy (§2). The philosophies of Plato and Aristotle, as well as of other ancient Greek thinkers continued to inspire philosophers in later times, even though critical views were also voiced at crucial moments in time.

A major rupture emerged in the 19th and 20th century, and that is the focus of this chapter. What was considered to be traditional metaphysical thinking was profoundly criticized by the German philosopher Friedrich Nietzsche (§3). We shall see that Martin Heidegger subsequently tried to depart from traditional metaphysics by developing an alternative (§4 and 5). Many philosophers have followed Heidegger in this move. Some of them believe however that Heidegger was still overly influenced by a metaphysical style of thinking. As a result, several critics of metaphysics ‘accuse’ each other of ‘still being too metaphysical’ (§6). In conclusion, it will be argued that much of the contemporary critique of metaphysics can be seen as a sort of metaphysics in itself, which makes metaphysics today a somewhat hyper-critical and self-undermining discipline, which has been declared dead many times and yet rises again each time (§7).

## 2. GREEK METAPHYSICS

From its very start philosophy has been looking for a stable and lasting understanding of being. It should come as no surprise, then, that philosophers were trying to find that which is of lasting value in reality. What is 'really' true is supposed not to be true for just a while, but rather forever. In this way, many Greek philosophers argue that rational or scientific knowledge concerns that which always stays the same and does not change. Continuity and sameness are therefore valued higher than change and variation, and likewise identity is valued higher than difference.

Given this focus, real knowledge is thought to be about reality 'in general'. It is about the invariable essences that underlie or shape our everyday experiences rather than about the variations in and changes of those experiences themselves in our daily lives. We can see many individual horses but scientific knowledge is about horses in general, that is, it is about the essential features that make a horse 'horse.' All individual, concrete horses share this general, abstract essence of 'being a horse'. In general, essences like 'horse-ness' are more fundamental than individual or singular things that actually exist like this specific horse in front of me.

The many horses that exist, that are born, live and pass away, are all captured by thought and knowledge in the one concept 'horse', which remains the same. Moreover, 'first philosophy' is not about all the different essences considered in isolation from each other, but it examines how they all hang together. Reality is not chaotic, so it is assumed, but rather reality is a coherent unity, and an examination of this unity is more fundamental than a study of the plurality of things. In short, in metaphysics we can discern the primacy of identity and stability over difference and movement, of the general over the singular, and of unity over plurality. The most important question then will be: what are the main characteristics of this unity, that is, of reality as a whole?

Given this object of investigation, the next question is *how* it is to be examined. Here the Greek philosophers contend that it is reason that can provide us with fundamental insights. They expect to find the principles of reality that underlie our experiences by employing rational reflection. This focus on reason establishes another 'primacy' in Greek philosophy: the primacy of reason over the senses. For the senses give us knowledge of specific things, while reason provides us with insight in essences. The assumption thereby is that reality itself is rationally ordered. Since reflection is a rational activity, this position can be described in the words of the early Greek philosopher Parmenides: "Thinking and being are the same" (Kirk e.a. 1983, 246n2).

The identification of reason and reality has had a profound influence in the history of philosophy, and is known as the 'identity-thesis'. One of the most famous formulations of it comes from the German philosopher G.E. Leibniz: *nihil est sine ratione*, nothing is without reason, or: nothing is without ground. Leibniz called this the 'principle of sufficient reason', and later philosophers have taken it to imply that in principle everything can be understood and explained by rational reflection and research. (Leibniz 1969, 227)

All of these primacies can be found in combination in the thought of one of the most important and influential philosophers ever, Plato. According to his famous theory of Ideas, real knowledge is not about singular things like, e.g., horses, but about the general form or 'Idea' of the horse. With our senses we may gather knowledge of several individual horses, but lasting knowledge can only be attained by rational reflection, by reason. Reason focuses on what all horses share, i.e. 'horse-ness' or the Idea of the 'horse.' According to Plato, a horse is a horse because it 'participates' in the Idea or the essence 'horse'. Behind the everyday world that we can see, hear, or touch, and that can offer us only temporary or even illusory knowledge, there is a world of essences, of Ideas, that we can come to know by rational insight. This is illustrated by his famous allegory of the cave (Plato, *Republic*, 514a-521b).

Though indebted to Plato, his pupil Aristotle is more interested in the reality as we observe it. In what was later called his *Metaphysics*, Aristotle discusses two questions: What are the principal qualities of a being? And, what is the highest being, that is, what is the principle that guides and arranges all other beings? He criticized Plato's assertion that only the Ideas are truly real. Instead, Aristotle argues that concrete sensible things are real. They are combinations of form (the abstract essence of something, what stays the same) and matter (the concrete material from which something is made, what changes over time), that develop towards their goal. The form 'horsehood' is what is essential to a horse, what makes this horse a horse, and ensures its identity over time: Lady the filly is the same horse (i.e. has the same form) as Lady the old mare, even if its matter changes (Lady's hair, skin and other cells will have changed over time). As one already said in Antiquity, Aristotle 'brought down' the Platonic forms to earth and located them in the things themselves. Since reality is still seen as forming a unity, all movements ultimately find their last goal in an Unmoved Mover, the highest being, which is a form of pure contemplation.

Plato's and Aristotle's views constitute two major metaphysical systems of ancient philosophy. In the long history of philosophy that followed, many metaphysical systems were proposed in the search for the ultimate foundation of reality: God, Subject, Consciousness, Spirit, History, etc.

But metaphysics has also always had its fair share of critics. The sceptic Pyrrho, for instance, denied the possibility of reliable knowledge and thought we should 'suspend' our judgments, rejecting not only metaphysical claims but any kind of dogmatic claim to knowledge. The history of metaphysics also includes philosophers who criticize the possibility of obtaining knowledge of the basic or ultimate structure of reality or who consider the whole enterprise as building castles in the air. So metaphysics not only comprises the study of some fundamental questions that gave rise to famous philosophical systems and the study of these systems themselves, but also the discourses of their critique. Without discussing the long history of metaphysics, we will now turn to this critique in more recent times.

### 3. RADICAL CRITIQUE OF METAPHYSICS: FRIEDRICH NIETZSCHE

A staunch critic of the metaphysical tradition is Friedrich Nietzsche, who lived in the second half of the 19th century. In his view, metaphysics is nothing but the misleading invention of a *Hinterwelt*, a world of 'Ideas' beyond the world as we know it. But, according to Nietzsche, such a 'world beyond' or 'world behind' is simply a misunderstanding: it is a product of the philosopher's imagination. These metaphysicians are seduced by language. Because we talk about horses as belonging to one kind, the suggestion arises that there 'is' something like *the* horse in general. There is, however, no reason at all to think that such an idea of horse contains an objective identity of all horses on some higher level, nor an objective validity over all horses (Nietzsche 2002, 20).

Similarly, Nietzsche criticizes the whole western philosophical tradition. In his eyes, not only metaphysics, but also Christianity and all traditional moral values are fundamentally mistaken – a view summarized in his adage "God is dead." (Nietzsche 2001, 119-120) Being more radical than his skeptical predecessors, he questions the principle of reason itself. According to Nietzsche, there is no solid rational order in reality, there only is the dynamic and ever-changing world that we know through our senses. Rather than trying to refute the rational argumentation for different metaphysical worldviews, Nietzsche sets out to describe where these views come from and how they could have developed. He does not just argue against metaphysics and Christianity as such, he explains their origins through a 'genealogical method' that unearths their historical roots (Nietzsche 2006).

For Nietzsche, not only metaphysics but all of our knowledge is finite and historically embedded and can therefore never arrive at eternal truths. Human knowledge can only be developed from cultural and historical perspectives. It is impossible to transcend such a perspective and view reality from a God's eye



view. "There are only interpretations," Nietzsche concisely stipulates (Nietzsche 2003, 139). This means there can be no beliefs about which we can be absolutely sure and no all-embracing or all-determining metaphysical principles that can be known. In Nietzsche's own view, the *Hinterwelt* is an illusion and the world is nothing but a chaos of competing forces that all have a 'will to power.'

The result of this radical criticism is often seen as a form of nihilism, the idea that there are no intrinsic values. Nietzsche, however, has his own approach to nihilism. In his view, the alleged truths of metaphysics, as well as the norms and values of Christianity and conventional morality were all part of the imagined *Hinterwelt*. All efforts to reach and understand this 'world behind' have resulted in the conclusion that the whole idea was an illusion from the start. It turned out to be worthless in itself and inevitably resulted in nihilism (Nietzsche 1967). The conclusion that there is no 'real world behind,' that there are no ultimate foundations or principles, is labeled by Nietzsche as 'negative nihilism.' His plea instead is for a 'positive nihilism,' one in which individuals can trust their own power and insights, being also willing to revalue all traditional values in order to create their own values. This includes embracing even the fatal misfortunes that inevitably will happen to us. Nietzsche thus tries to replace the perspectives of truth versus untruth, and good versus bad with a 'transvaluation of all values' (Nietzsche 1918, 182). He criticizes all traditional, rationalistic, and Christian values as being depressing and unhealthy, claiming that it is healthier to develop your own strength and your own values.

In short, Nietzsche radically rejects all characteristics of traditional Metaphysics: the primacy of generality, unity and identity over singularity, plurality and difference, as well as the identity thesis and the principle of reason. This criticism has been very influential. But did it really mean the end of Metaphysics?

#### **4. OVERCOMING METAPHYSICS: MARTIN HEIDEGGER (1)**

In reaction to this vigorous attack on the metaphysical tradition, we can see a parting of two ways in twentieth century philosophy. The first of these two, 'analytical philosophy', starts with the call to restrict philosophy to explaining and analyzing meaningful statements made in science and philosophy. Traditional metaphysical claims were considered to be literally meaning-less: they could not be verified or falsified, unlike (most) scientific statements, and hence no meaning could be attached to them. Among philosophers known as the logical positivists (the Vienna Circle and its followers), metaphysics had a bad reputation. For much of the twentieth century analytical philosophers shared this negative view of traditional metaphysics, and there was hardly any contact between them and the second current of thought, 'continental philosophy'. Only in more recent times, a fruitful interaction between analytical and continental

philosophers has begun to emerge, even though differences in style and content remain. In addition, metaphysical questions have recently received their own treatment again from analytical philosophers.

Whereas analytical philosophers often simply ignored traditional metaphysics, continental philosophers (broadly speaking, those philosophers working in Germany and France) criticized it by seriously engaging with it. They asked that attention be paid to that which appears to escape or exceed human reason, rational reflection, and conceptual thinking. This critique of metaphysics is certainly not less radical, but focuses on different aspects of the metaphysical tradition. One of the most influential voices in this debate is Martin Heidegger, a twentieth century German philosopher, who tried to overcome traditional metaphysics by developing an alternative for it.

Heidegger starts with the question of Being, which for him is the most fundamental question in philosophy. What does it mean that things 'are', that there is a world, that we exist? It is not so easy, however, to ask the question of Being in a proper way. For the question 'what is Being?' in itself already uses the word 'is,' a conjugation of the verb to be, and therefore already presupposes a view of Being, one that should be questioned instead of implicitly affirmed. Hence, 'Being' cannot simply be defined in a phrase like "Being is...". (Heidegger 1962, 21-64)

According to Heidegger, from the very beginning of philosophy, Being has been misunderstood. Despite a promising start by the pre-Socratics, philosophers never thoroughly realized the difference between Being and beings. Beings, regardless whether they are things like cars and buildings or occurrences like traffic or art performances, have one 'thing' in common: they 'are.' But what this 'are' means cannot itself be understood as a being, as existing as a thing or occurrence. The fact that beings exist, and the way they exist, 'is' not itself a being. In English, this is usually expressed as the Being (*Sein*) of beings (*Seiendes*), or the difference between Being and a being. Heidegger calls this difference between Being and beings the 'ontological difference'.

In Heidegger's view a major mistake was made in the beginning of philosophy by confusing Being and beings. From the start, the basic principles of reality were thought of as beings. Aristotle's metaphysics indicates how this manner of thinking results in two main questions: What are the main characteristics of a being? and, What is the highest being? This last question necessarily arose as soon as one started to think in terms of beings, because it was thought that the orderly arrangement of beings can be understood only on the basis of a principle, that is, the most fundamental being. Heidegger labels this way of thinking as 'onto-theology', taking the two metaphysical topics of Aristotle together as

ontology and theology (Heidegger 2002a). The traditional focus on beings sadly implied the “forgetfulness of Being” (Heidegger 1962, 21-35).

But how, then, can we understand this ‘Being’, if it is not a being or thing? This is not an easy question to answer. In all his reflections, Heidegger never provides a clear and unequivocal description of Being. A hint in the right direction comes from language: the word ‘Being’ should first of all be understood as a verb and not as a noun. The reality we live in, is dynamic and versatile, always in movement and change. Despite the constancy and order that we also see around us, it is too dynamic to be fixated in solid concepts. What is, happens, and can only partially be understood. In developing knowledge of the world around us, we must fix and determine what actually is in motion all the time.

In line with Nietzsche, Heidegger states that all knowledge, including all of experience, is a matter of interpretation. It is all dependent on specific practical and cultural perspectives, by which some elements of what we can know may appear, whilst others remain concealed. We can consider a horse to be large, expensive, beautiful, our possession, powerful, healthy, fast, working hard, elegant, black, useful for an escape or representative for its race, but we cannot have all these experiences at the same time. Given that things only appear to us under these conditions, Heidegger does not defend relativism, because that still presupposes the possibility to compare perspectives from a God’s eye view. He claims that the dilemma between either objective knowledge that would be absolutely certain, or subjective relativism, needs to be overcome. We can have reliable knowledge, but it is always dependent on specific perspectives and circumstances.

In every appearance of beings some features remain concealed. Being only manifests itself implicitly in the beings that appear to us and hence remains partly concealed itself. The same applies to the dynamic nature of Being, which cannot be grasped in the fixed and stable concepts that we employ. Given these limitations, Heidegger searches for a new understanding of Being. It needs to be thought as a dynamic occurrence of beings, not as a being that would manifest itself as a firm foundation or a guiding principle.

The entire history of philosophy, however, has been dominated, at least in Heidegger’s view, by the metaphysical effort to find a highest principle that is the source, foundation or goal of everything. Not only Plato’s Ideas, Aristotle’s Unmoved Mover and the Christian God, but even Nietzsche’s Will to Power, are all manifestations of this metaphysical way of thinking. These highest principles all lay the basis for an assembling of reality that is ordered by hierarchical oppositions like necessary and contingent, body and soul, high and low, reason and nature, subject and object, and so on. Therefore, in Heidegger’s view, metaphysics itself

is also a manner of thinking that tries to grasp and control reality by forcing it into hierarchically organized systems.

In this respect, Heidegger regards Nietzsche's philosophy and the modern era of technology as the last phases of a metaphysical tradition that has always focused on beings as things that can be defined, explained, calculated, and controlled. Coming at the end of this metaphysical tradition, Heidegger believes that we need to overcome it and look for a new beginning.

He calls this project the *Destruktion* of the history of ontology. This should not be translated literally as 'destruction', but as dismantling. This re-interpretation of the historical metaphysical systems is aimed at laying bare their inner contradictions and blind spots. Such a dismantling is meant to pave the way for a new style of thought (Heidegger 1962, 41-49). Heidegger even speaks of 'the end of philosophy and the task of thinking', suggesting that philosophy needs to be replaced by an entire new manner of thinking (Heidegger 2002b).

## 5. A NEW BEGINNING FOR PHILOSOPHY: MARTIN HEIDEGGER (2)

In several ways Heidegger searches for a new beginning of philosophy. He interprets the fragments of the earliest Greek philosophers such as Anaximander and Parmenides, in search of traces of another non-metaphysical thought. In need of a new language or a re-invention of existing language, he also tries to find inspiration in poetry, mainly in the work of the German poet Friedrich Hölderlin. The philosophical terminology is too contaminated by what he calls "the language of metaphysics." It contains terms like 'subject,' 'essence' and other words that could not but lead the thinker to assume or defend a fixed and allegedly stable system of thought. However, Heidegger also states that overcoming metaphysics is an endless task, because one inevitably falls back to the language of metaphysics. Probably, the 'seduction of language,' which was an expression of Nietzsche, is inevitable.

In contrast to the metaphysical tradition that he seeks to overcome, Heidegger's thinking shows a primacy of difference over identity, of the singular over generality, and of plurality over unity. Heidegger also explicitly distances himself from the principle of reason. He re-interprets this principle – *nihil est sine ratione*, *Nichts ist ohne Grund*, nothing is without sufficient reason – in such a way that in the end it means something like the opposite. Through "Nothing' is: without ground" Heidegger arrives at: "Nothing: the abyss." (Heidegger 1991). Thus, according to Heidegger, instead of a metaphysical first principle, the abyss of a "nothingness" can be found behind all things. However, he also tries to avoid Nietzsche's nihilism. Rational reflection can give us helpful insights in reality but it cannot give us absolutely certain knowledge or a survey of the world as a whole. Hence,

without leaving reason behind, Heidegger disagrees with the metaphysical presupposition that reality can in principle be completely comprehended by rational reflection.

Heidegger's philosophy clearly embodies a radical critique of the metaphysical tradition. His new alternative way of thinking, however, remains an endless effort of searching and questioning. The overcoming of metaphysics can never really succeed; it can only continue as an infinite task. Moreover, it is not only a task, but also an unclear destiny. Heidegger's later thought emphasizes that human thinking is profoundly guided by long term developments that it cannot survey or choose. Our thoughts are not created by ourselves, but mainly come to us from the 'history of Being'. Within this history of thought, the era of metaphysics seems to come to an end and philosophy is in need of a new beginning, but it is not really in our power to develop a new way of thinking.

In the end, therefore, it is not clear whether and to what extent metaphysics has actually ended or whether it has been overcome. Did metaphysics really come to an end? Or was Heidegger's own thought still marked by metaphysical traits? Is it ever possible for philosophy to engage in systematic reflection that is not reminiscent of the metaphysical effort to reach a solid foundation or principle? Philosophers who have followed in Heidegger's footsteps have given different answers to these questions.

## 6. AFTER METAPHYSICS OR CONTINUING METAPHYSICS?

**D**espite Heidegger's efforts to overcome metaphysics, the twentieth-century French philosopher Jacques Derrida claims to trace metaphysical traits in Heidegger's own work. On the one hand, and in line with Heidegger, Derrida criticizes metaphysics insofar as it aims to construct ordered views of reality on the basis of a foundation, source, or goal. He deconstructs or dismantles these constructions by showing how they, with their alleged foundations, are always embedded in social and linguistic networks of references that are, by their very nature, unstable. At the same time, Derrida regards such constructions as unifying systems that reduce all differences and 'otherness' to the same general concepts, within hierarchical oppositions (Derrida 1976).

On the other hand, Derrida also states that this metaphysical way of thinking is inevitable. We cannot but think with a general vocabulary that reduces differences to identities. The language of metaphysics is therefore unavoidable. Even the conception of and clear distinction between two periods and two styles of thought – metaphysics and beyond metaphysics – would in itself be a clear opposition and thus a sign of metaphysical thinking (Derrida 1988; 1979, 117-119).

In this view, philosophical and scientific research cannot but result in constructions of knowledge that may be practical and helpful, but that also need to be dismantled in order to open up new and unexpected perspectives. Derrida tends to see in every construction the contours of a metaphysical system, and such a system needs to be deconstructed. What seems to be left, then, for philosophy, is an endless dismantling of inevitable metaphysical constructions.

Political or juridical reforms, for instance, need to aim at justice, but because no one can give a final definition of justice, it infinitely remains open to improvement. To give another example: Derrida criticizes the attempts to draw a fundamental distinction between human beings and 'animals,' because such an attempt takes many different animals together as categorized under the same 'essence,' namely, being a non-human animal.

Following another aspect of Heidegger's inheritance, the German philosopher Hans-Georg Gadamer presents a different elaboration of Heidegger's notion of *Destruktion*. He does not aim to deconstruct an inevitable metaphysics, but tries to lead strict philosophical terminology back to its origins in living dialogue, where the meaning of language is never entirely fixed. According to Gadamer, it is possible to think without a language of metaphysics. Philosophical reflection can do without an ultimate principle or foundation. The philosophies of difference that keep on fighting against the metaphysical tradition, in fact prolong it in a negative way (Gadamer 2007). In a similar fashion, another German philosopher, Jürgen Habermas, has argued for a post-metaphysical philosophy that stops asking for ultimate principles (Habermas 1992).

The apparent paradox of this discussion is that these French and German philosophers accuse each other of still being too metaphysical, as if this would be a crime or a disease. In fact, they use the term 'metaphysics' with different meanings. According to Derrida, thinking with clear concepts and distinctions always uses a metaphysical style and language, but this language, in spite of its appearance, is in itself unstable and refers to something unstable, which is the task of the philosopher to disclose. In addition, this metaphysical language is problematic, because it reduces unique otherness to general frameworks, which are based on absolute principles as Ideas, God, etc. Metaphysics is an inevitable but problematic feature of philosophy that has to be dismantled again and again.

According to Gadamer and Habermas, such criticism, however, is still based on the misleading presupposition that philosophy cannot but think in terms of pure distinctions, based on absolute principles, which then need to be undermined and dismantled. Not every distinction between, for instance, justice and injustice, or humans and other animals, should be reduced to alleged general metaphysical systems. In their view, only the explicit search for absolute principles should be

called 'metaphysics'. Philosophical reflections that reveal awareness of their historical and cultural contexts can avoid such broad questions and claims.

From Derrida's perspective, however, such more modest reflections can only be offered within a general framework that necessarily builds on metaphysical presuppositions. Therefore, metaphysical questions are unavoidable after all. According to Gadamer, this may be true but it does not make it necessary to explicitly discuss these questions on a metaphysical level. In Gadamer's view, Derrida's approach of criticizing and dismantling systems of thought only works because it presupposes such systems of thought in the first place, as if philosophy always reduces multiplicity and difference to universal systems; this presupposition, Gadamer says, we don't need to accept.

On closer inspection, both approaches – that of Derrida on the one hand, and Gadamer and Habermas on the other – focus on different aspects of the metaphysical tradition. One may see metaphysics as the domain of the big questions and all-surveying systems or as reductions of unique and dynamic life to fixed frameworks of conceptual constructions with hierarchical oppositions, for instance, the reduction of all non-human living creatures to 'animals,' or implicitly using the concept and features of 'man' as a model for 'human.' According to Derrida these two sides of metaphysics are in fact always related. In the eyes of Gadamer and Habermas, making a conceptual distinction or mutual understanding is not the same as a reduction of otherness to general rational concepts, and it also does not necessarily include large-scale conceptual systems of reality as a whole. Only the last aspect would be labeled as 'metaphysics' by Gadamer and Habermas. Nevertheless, their views on language and knowledge, including their conditions and possibilities, may be regarded as points of view on a metaphysical level because they contain presuppositions about the limits and possibilities of human knowledge as such.

## 7. CONCLUSION (TO BE CONTINUED)

**T**hrough the ages, metaphysics has often been regarded as the most fundamental domain of philosophy, which inquires after the ultimate or foundational principles of reality. At the same time it has always attracted a lot of criticism both in Antiquity and in later periods. From the nineteenth century onwards, metaphysics began to be criticized so vehemently that it was often declared dead. The world as a whole was considered too large to be surveyed and comprehended by human reflection, which is inevitably conditioned by its own perspectives. In addition, life and the world were seen as too vivid, unique, and inconstant to be grasped by a set of basic principles.

But pleas for a complete abandonment of metaphysics were not successful, because the metaphysical questions about the unity and main characteristics of reality, as well as about the most basic presuppositions of other disciplines, returned again and again. Martin Heidegger attempted to develop an entirely new style of thinking, different from the metaphysical tradition, which he called a new 'thinking of Being.' But it appeared to be very difficult to throw away the 'language of metaphysics' with its general concepts, basic principles, and fixed terminology. Such efforts to replace metaphysics have turned out to be more like a renewal of metaphysics. The many critiques of metaphysics can be considered as metaphysical reflections in a new manner, continuing metaphysics by criticizing it.

Today many philosophers prefer to avoid these kinds of big questions, are not interested in such questions, or think they don't make sense. They study moral questions, the differences between human beings and other species, the concept of truth, knowledge, and all the other themes addressed in this handbook. Indeed, as we already noticed, such a turn to and specialization in different subdomains is an important feature of contemporary philosophy and of the sciences in general. Yet insofar as those different endeavors endure, we still assume that there is a criterion of right and wrong, that there is a conception of truth, or a clear demarcation between humans and other creatures, and so on. As such, we may be said to be doing metaphysics after all: we always can, and every now and then have to critically question, on a metaphysical level, the implicit presuppositions of the very concepts that we use. Examining life cannot be done without such very broad-scale questions about, for instance, truth, ethics, and life. Therefore, despite having been burned down to the ground many times, metaphysics, the discipline of the most fundamental questions, always rises again from its ashes.



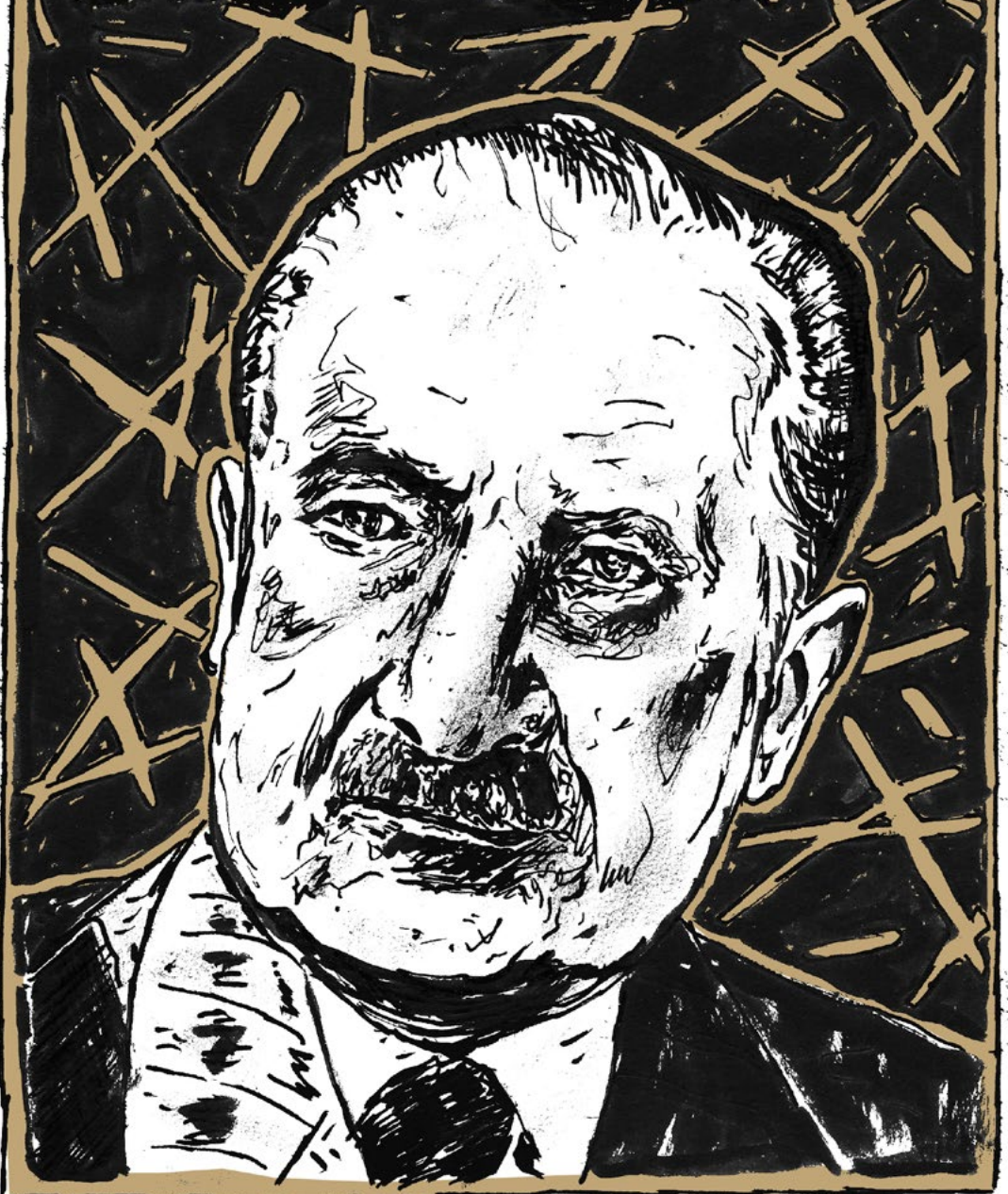
## BIBLIOGRAPHY

- Aristotle (1984). *Metaphysics*. In J. Barnes (Ed.), *The complete works of Aristotle II* (pp. 1552–1728). Princeton University Press. (Original work published 4th century B.C.E.)
- Carnap, R. (2003). *The logical structure of the world and pseudoproblems in philosophy*. Open Court.
- Derrida, J. (1976). *Of Grammatology* (G. C. Spivak, Trans.). John Hopkins University Press. (Original work published 1967)
- Derrida, J. (1979). *Spurs: Nietzsche's Styles*. University of Chicago Press.
- Derrida, J. (1988). *Limited Inc.* Northwestern University Press.
- Dilthey, W. (1989). Concluding observations concerning the impossibility of a metaphysical approach to knowledge. In *Selected Works I: Introduction to the human sciences* (pp. 219–240). Princeton University Press.
- Gadamer, H. (2007). Heidegger and the language of metaphysics. In R. E. Palmer (Ed.), *The gadamer reader. A bouquet of the later writings* (pp. 345–355). Northwestern University Press.
- Habermas, J. (1992). *Postmetaphysical thinking* (W. M. Hohengarten, Trans.). MIT Press.
- Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Blackwell. (Original work published 1927)
- Heidegger, M. (1991). *The principle of reason* (R. Lilly, Trans.). Routledge. (Original work published 1957)
- Heidegger, M. (2002a). The end of philosophy and the task of thinking. In J. Stambaugh (Trans.), *On time and being* (pp. 55–73). Chicago University Press. (Original work published 1957)
- Heidegger, M. (2002b). The Onto-theo-logical constitution of Metaphysics. In J. Stambaugh (Trans.), *Identity and difference* (2nd ed., pp. 42–74). Chicago University Press. (Original work published 1957)
- Kant, I. (1999). *Critique of pure reason*. Cambridge University Press. (Original work published 1781)
- Kirk, G., Raven, J., & Schofield, M. (1983). *The presocratic philosophers*. Cambridge University Press.
- Leibniz, G. W. (1989). The Principles of Nature and of Grace, Based on Reason. In L. E. Loemker (Ed. & Trans.), *Philosophical Papers and Letters* (pp. 636–642). Springer Netherlands. [https://doi.org/10.1007/978-94-010-1426-7\\_67](https://doi.org/10.1007/978-94-010-1426-7_67)
- Nietzsche, F. (1918). *The Antichrist* (H. L. Mencken, Trans.). Knopf.
- Nietzsche, F. (1967). *The will to power* (W. Kaufman, Trans.). Vintage Books.
- Nietzsche, F. (2001). *The gay science* (A. Del Caro, Trans.). Cambridge University Press. (Original work published 1882)
- Nietzsche, F. (2002). *Beyond good and evil* (J. Norman, Trans.). Cambridge University Press. (Original work published 1886)

Nietzsche, F. (2006). *On the genealogy of morality* (C. Diethe, Trans.). Cambridge University Press. (Original work published 1887)

Plato. (1935). *Republic* (P. Shorey, Trans.). Harvard University Press.

MARTIN HEIDEGGER



# MARTIN HEIDEGGER

## SEIN UND ZEIT / BEING AND TIME

---

**M**artin Heidegger's *Being and Time* is dedicated to the question of Being, according to Heidegger the main question of all philosophy. In the philosophical tradition, Being was implicitly understood, in terms of time, as a form of presence: being here and now. Past and future were secondary modes of Being and time, showing a lack of presence: what is in the past, is no more and what is in the future is not yet. The highest being, therefore, presupposed by all beings, was assumed to be a being that is always here and now, an eternal divine being. In contrast, Heidegger is looking for a more dynamic understanding of Being and time. It is difficult, however, to articulate the question of Being in the right way. One cannot ask for a definition, because that would already presuppose the use of the verb 'to be' and thus an idea of Being. Moreover, a definition fixates and delimits, while Being is supposed to 'be' dynamic and all-embracing. In fact, *Being and Time* is nothing but the preparation for a new way of asking the question of Being. In profound and precise analyses Heidegger describes the main characteristics of the being that is able to ask the question of Being: by this being Heidegger means us, human beings. *Being and Time* can therefore also be read as an anthropological study of human existence. The book was first published as the first volume in a series of two or three volumes, but the other volumes were never written. Neither did Heidegger succeed in finding the right language to ask the question of Being. Nevertheless, *Being and Time* is one of the most important books of 20th century philosophy, a landmark that has set the stage for later developments in contemporary philosophy.

---

Menno Lievers

**Language and Reality**

# LANGUAGE AND REALITY

## *An Introduction to Philosophy of Language*

Menno Lievers  
[M.Lievers@uu.nl](mailto:M.Lievers@uu.nl)

### 1. INTRODUCTION

In the 20th century philosophy has taken a 'linguistic turn'.<sup>1</sup> Philosophical progress, so it was thought, could only be achieved through an analysis of the meaning of the words in which philosophical problems were formulated. The linguistic turn was motivated by two tenets: the first is that the only way to analyze thought is via an analysis of language. The second tenet is that *our* world is structured by the language we speak. As a result, philosophers in the analytical tradition dedicated their efforts mainly, if not exclusively to language. This examination was a means to an end; their interests were not in language itself, but in how an analysis of language could solve philosophical problems. We could call this 'philosophy of language' in the broad sense, by which is meant a particular way of practising philosophy, a philosophical methodology. Some philosophers even went as far as stating that the philosophy of language had become *prima philosophia* in Aristotle's sense, by which is meant that prior to addressing problems in other areas of philosophy, issues within the philosophy of language have to be resolved. Even the ontological debate about realism (the view that there is a structured, ready-made world out there awaiting discovery) and idealism (the view that the structure of the world is somehow dependent on the mind) had to be reformulated and settled in terms of the philosophy of

---

1. "The linguistic turn" is a collection of essays edited by Richard Rorty and published in 1967. Rorty attributed the phrase to Gustav Bergmann. See R. Rorty (1991).

language. This approach is to be distinguished from 'philosophy of language' in a narrow sense, a branch of philosophy, closely affiliated to linguistics, that aims to describe and explain features of language that are of interest in themselves.

Given their underlying motivation, philosophers of language in the 20th century focussed on the relationship between language and the world, and language and the mind. The central problem in this regard was: what is meaning? A theory of meaning was supposed to give an account of the relationship between language (and thus the mind) and the world. 'The reference relation' is crucial in this regard: does the meaning of words determine what objects they refer to in external reality, which would be congenial to idealists, or are words like labels that can be attached to objects that are mind-independent items, as realists would have it? Does language actively shape our view of the world, or is there a world independent of our linguistic characterization?

There are two conflicting intuitions at work while attempting to answer the question what meaning is. The first is that speakers of a language determine themselves what their words mean. In the history of philosophy this *mentalist* intuition is known in particular of the account by John Locke (1632–1704), according to which a word uttered by a speaker stands for (or refers to) to the corresponding idea in the mind of that speaker. When I use the word 'dog', it can stand only for the idea dog in my mind, not to yours nor to the dog itself (only indirectly via the idea). (*Essay* 4.4.3) To avoid the suggestion that there is a necessary connection between a word and an idea, Locke insists on the 'inviolable right' I have to assign my own word to my own idea (3.2.8). But if every speaker is the arbiter of the meaning of his or her own words, communication becomes impossible. Locke admits that we have to conform ourselves to the common usage of a word on pain of misunderstanding, and indeed common language 'regulates the meaning of words pretty well for common conversation' (3.9.8), but the immediate meaning of the word, he insists, can only be the idea in the speaker's mind.

The mentalistic thesis of words standing for mental ideas invites a simple objection that is based on the phenomenology of linguistic understanding. If we hear someone speak and if we understand what he or she is saying, we hear immediately meaning in the sounds they utter and do not have to figure out via a process of calibration and interpretation what the speaker wants to convey.

The opposite intuition is that the meaning of language resides in a relationship between language and the world. The simplest expression of this intuition is the naïve theory of meaning,<sup>2</sup> according to which the meaning of a proper name is

---

2. In Wittgenstein's *Philosophical Investigations* this theory is called 'the Augustinian picture of language'. Ryle calls it the "Fido" – Fido principle". Ryle (1949), pp. 69-70.

the object or person it stands for.<sup>3</sup> The naïve theory of meaning presupposes that there is a 'ready-made' world out there, to which words can be attached like labels; it thus presupposes ontological realism. Many philosophers in the 20th century felt that traditional questions, like the controversy between realism and idealism about the external world, could be reformulated in terms of the philosophy of language. In this chapter I describe how the development of the philosophy of language in the 20th century is determined by the struggle between these two intuitions. I also examine whether philosophers of language subscribe to realism or idealism.

## 2. FREGE

**D**espite its intuitive plausibility the naïve theory of meaning faces objections that are hard to overcome. The German philosopher and mathematician Gottlob Frege (1848–1925) raised two objections in his seminal article "On Sense and Reference" (1892). The first is known as the problem of the informativity of identity statements.<sup>4</sup> During a certain part of the year a heavenly body is visible during the morning and has accordingly been called 'Morning Star'. A few months later the morning star has disappeared, but a heavenly body is visible during the evening that just as appropriately has been called 'Evening Star'. According to lore, Pythagoras discovered that the planet Venus that was called 'Morning Star' during part of the year, was the same heavenly body that had been called 'Evening Star' during another part of the year. One thus discovered that the identity-statement 'Morning Star is Evening Star' is true in reality. One knew the meaning of the words 'Morning Star' and 'Evening Star', but one did not know that these were two names for the same planet.

This constitutes a problem for the naïve theory of meaning. The reason is simple: if the meaning of a noun is the object the name stands for, then the 'Morning Star' has the same meaning as the 'Evening Star', namely the planet Venus. Someone who knows the meaning of these nouns, should therefore, without examining the heavenly sky in reality, be able to know that the identity-statement 'Morning Star is Evening Star' is true. That identity-statement should then be just as uninformative as tautologies as 'Morning Star is Morning Star' or 'The Evening Star is Evening Star'. For short: the naïve theory of meaning cannot explain why identity-statements such as 'Morning Star is Evening Star' are informative.

Frege offered a second objection. The Dutch Marxist Henk Sneevliet was well acquainted with Leon Trotsky. We are entitled to say that 'Sneevliet knew that

---

3. The theory gives a similar account for predicates, which stand for properties, and sentences, which stand for states of affairs.

4. Also known as the 'co-reference problem'.



Trotsky was the architect of the Red Army.' After having read his biography we know that Trotsky was born as 'David Bronstein', a fact Trotsky tried to conceal. If the naïve theory of meaning is correct, the names 'Bronstein' and 'Trotsky' have the same meaning. If that were true, we ought to be able to substitute 'Bronstein' for 'Trotsky' getting 'Sneevliet knew that Bronstein was the architect of the Red Army.' However, it is likely that Sneevliet did not know this. So, we are *not* allowed to substitute 'Bronstein' for 'Trotsky', since the substitution does not preserve the truth-value of the initial statement. Once again, we have to conclude that there is more to meaning than just the object the word stands for.

Frege raised two further objections. First, people make up stories and, with some exceptions, we are able to make sense of them. But how are we able to do that, if the naïve theory of meaning is correct? Names in fiction do not stand for objects in reality, so they cannot be meaningful. But they are. So, the naïve theory of meaning cannot be correct. The next objection is closely related to the previous one: we occasionally assert that something or somebody does not exist. 'Santa Claus does not exist' is such a negative existential statement. But how do we understand these statements, according to the naïve theory of meaning? Once again, we have to conclude that there must be more to meaning than just the object the proper name stands for.

Frege provided the following solution. When we talk about the planet Venus using the proper name 'Evening Star', we are not only talking about the heavenly body, but we are also presenting and thinking about that object in a particular way. We can talk about a countless number of things by using different names or descriptions for them. Frege presents the example of the intersection of the medians of a triangle. That point can be referred to as 'the intersection of median *a* and median *b*', but also as 'the intersection of median *a* and median *c*', or as 'the intersection of median *b* and median *c*'. The very same intersection-point is thus presented to us in three different ways. This *mode of presentation* is, according to Frege, an important ingredient of the meaning of a word. It is this mode of presentation that distinguishes the meaning of the proper name 'Evening Star' from the meaning of the proper name 'Morning Star'. Frege, therefore, concluded that the meaning of proper names and of words in general not only consists in that what words stands for (or designate, refer to) in reality, but also in the *sense* that contains the mode of presentation of what the words stand for.

With respect to the controversy between realists and idealists Frege's philosophy of language combines insights from both sides of the debate. On the one hand he concedes to the idealists that we need to grasp the meaning (the sense) of the word Evening Star in order to recognize the planet Venus as the Evening Star. So, the way we perceive reality is determined by the language we speak. On the other hand, Frege assumes with the realists that reality is structured into discrete

objects. There is a planet Venus out there that we can perceive in a particular (Evening Star) kind of way.

Does this notion of sense indeed solve the problems concerning meaning?

### 3. RUSSELL'S THEORY OF DESCRIPTIONS

One of the first philosophers who acknowledged Frege's genius was Bertrand Russell. Despite his admiration he was also a stern critic.<sup>5</sup> In what has become a classic paper, 'On Denoting', Russell observed that Frege's semantical theory, and in particular his notion of sense, cannot account for the meaningfulness and truth of many ordinary sentences. For example, 'A man is walking in the street'. The sentence is true, there are several men walking in the street, but to whom does the noun 'a man' refer? Is the notion of sense (mode of presentation) of any help here?

Russell applied the technique of formal logic (see chapter "The Study of Informational Processes" of this handbook), which Frege had put forward in 1879, to show that the grammatical *surface structure* of a sentence is different from the *logical deep structure*. In the sentence 'A man walks in the street' 'a man' is the grammatical subject. The sense of 'a man' is, according to Frege, the mode of presentation of the reference. But to whom does 'a man refer'? And if there is no referent, how are we to decide who is walking in the street and thus whether the sentence is true or false? Russell solves this problem as follows: logical analysis reveals that what is being said in 'A man walks in the street' is 'There is at least one entity that is a man and that entity is walking in the street.'

A related difficulty Russell observed concerns so-called *definite descriptions*, like 'the present prime minister', which were regarded by Frege as being semantically similar to 'real' proper names, like 'Johnson' or 'Merkel'. If Frege is right, a sentence like 'The present King of France is bald' cannot be true or false, since the definite description 'The present King of France' does not refer to anybody. According to Russell, this is a mistake. He argued that this sentence is false. If Frege's philosophy of language is correct, the phrase 'the present King of France' must have an object to which the phrase refers to after all, otherwise the sentence cannot be true or false. Since in our actual world there is no King of France at present, some philosophers therefore assume that there is an abstract realm outside space and time in which fictional objects, like the present King of France, exist. Russell vehemently objected to this move, which he took to be a too large departure from common sense. Instead, he revised Frege's analysis of

---

5. Most notably Russell discovered the set theoretical paradox in Frege's work on the foundations of mathematics.

sentences. In Russell's analysis, the logical deep structure of the sentence about the king of France being bald is a conjunction of three propositions: There is at least one entity that is King of France & There is only one entity that is King of France & That entity is bald. Since the first conjunct is false, the entire sentence is false, which gives Russell the result he was looking for. Russell's common sense, which he called his 'robust sense of reality' was thus satisfied.

#### 4. ORDINARY LANGUAGE PHILOSOPHY

For decades Russell's theory of descriptions was hailed as 'that paradigm of philosophy'. It therefore came as a shock, certainly for Russell, that in 1950 P. F. Strawson published an article in which he argued that Russell had ignored important distinctions. 'The present King of France is bald' is a perfectly meaningful sentence in English, but no one in his or her right state of mind will presently (neither in 1950 nor in 2021) use that sentence to assert that the thought expressed in that sentence is true or false, because there is no King of France. A meaningful assertion of that sentence in real life *presupposes* that there is a King of France. According to Strawson, we ought to distinguish, first, *type sentences*, for instance as they occur in schoolbooks on grammar, from, second, uses of sentences in certain type circumstances, and third, *token utterances* by a particular speaker on a particular occasion. Only the last, token utterances, are eligible for truth and falsity. Type sentences do not refer; only speakers do in acts of asserting sentences.

Strawson's approach was representative of the way philosophy was practised in Oxford in the first decades after the Second World War. So called 'ordinary language philosophy' adhered to a distinctive philosophical methodology: many, if not all, philosophical problems would disappear, if only philosophers paid sufficient heed to the way the central notions in terms of which a philosophical problem was formulated were used in ordinary language.

Already in 1949 Gilbert Ryle had published *The Concept of Mind*. In the first chapter he famously analysed the Cartesian mind-body problem. According to Ryle, the Cartesian mind was a 'para-mechanical hypothesis', motivated by a wrong-headed theory of meaning. Ryle labelled this mistaken theory of meaning the "Fido" – Fido principle, his name for what we have called 'the naïve theory of meaning' (the name 'Fido' refers to the dog Fido). If the meaning of a word is the object it stands for, then, since the word 'mind' is meaningful, it ought to stand for an object as well. This, however, Ryle argued, is a *category mistake*. The mind is not a kind of thing on the same par as a body or an object. The example Ryle used to illustrate this view is the following. A visitor to Oxford wants to see the University of Oxford. He is being shown the colleges and the Bodleian Library. When the tour is over, the visitor inquires "But where is the University of Oxford?"

It has escaped his notice that the University of Oxford is an entity of a different category: it is the organization of the colleges, not an institution that requires a physical infrastructure. The visitor has already seen the University of Oxford. In the same manner an examination of the way in which words for mental states are used in daily life, e.g. 'being afraid' and 'anxious', establishes that they do not refer to mental states hidden behind someone's eyes, but instead to publicly observable, overt behaviour.

Analysing ordinary language could solve philosophical problems. One of the most prominent ordinary language philosophers, J. L. Austin, wrote: "[...]"[...] our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our armchairs of an afternoon—the most favoured alternative method."<sup>6</sup>

Although Austin's articles exhibit the ordinary language approach to philosophy, his main and most lasting contribution was to the study of language as an end in itself rather than as a goal clarifying and solving philosophical problems. His 1955 William James lectures *How to do things with words* were the impetus to the speech act theory, that inaugurated the rise of a branch in linguistics called pragmatics, which focuses on the use of language in interaction.<sup>7</sup>

In these lectures Austin criticized the tendency within the philosophy of language developed since Frege to focus almost exclusively on sentences that can be true or false. Austin pointed out that many linguistic utterances are not meant to be assertions that can be true or false, but acts that are governed by highly specific criteria of correctness. Making a promise, for instance, requires that the speaker is sincere, that he or she lives up to the promise and that what he or she promised can be achieved.

Once attention was being drawn to these aspects of language, it dawned rather quickly on philosophers that language use is often intended as a performative action. *We perform* or do something by speaking, e.g. by promising something. Subsequently, Austin came to the view that assertions are also acts, *speech acts*, just like promises, and that they ought to be regarded as a separate subclass.

---

6. Austin, J. L., (1956-7) [1961]. "A Plea for Excuses", p. 130.

7. Austin's student J. R. Searle ought to be mentioned in this regard as well, in particular his book *Speech Acts*, [1969] (1976).

Austin attempted to set up a systematic speech act theory. Any linguistic utterance is a speech act, in which we can formally distinguish how and which words are uttered in a syntactically orderly fashion (which he called 'the locutionary aspect'), what kind of speech act is performed ('the illocutionary aspect'), and what effect the speech act is supposed to establish ('the perlocutionary aspect'). For instance: if someone standing next to a frozen pond yells: 'The ice is thin!', the locutionary aspect is the syntactically correctly formulated indicative 'The ice is thin'; the illocutionary aspect is a warning, and the perlocutionary aspect is the effect on the audience of not attempting to walk on the ice.

Austin tried to give a systematic classification of speech acts on the basis of their illocutionary aspect. However, it turned out to be extremely difficult to provide a systematic taxonomy of speech acts in a principled manner. Some philosophers argue that this was to be expected, since any attempt to develop a systematic account of meaning ignores the context-sensitivity of language.

With respect to the question of the opposition between realism and idealism ordinary language philosophers are nowadays often viewed as being committed to a naïve realist, common sense view of reality, since that is the conception of reality that is presupposed by ordinary language. Opposed to Austin's reverence for ordinary language was his *collega proximus* at Oxford, A. J. Ayer, who believed, on the contrary, that ultimately all problems, even philosophical problems, were empirical ones that ultimately had to be solved by natural science.

## 5. LANGUAGE AND THE UNITY OF SCIENCE

Ayer's source of inspiration was the Vienna Circle. In the thirties he had visited Vienna and he had met the Circle's most prominent members, Schlick, Carnap, Neurath, who defended what they called 'logical positivism'. Impatient with the lack of progress in philosophy they urged that the scientific method ought to be applied in philosophy. They rejected intuitions as a reliable source of knowledge on the basis of the following argument. According to the logical positivists, the appeal to intuitions has led to meaningless metaphysics forming houses of cards that can be blown away if we adhere to the methodology of empirical science. In their battle against metaphysics the logical positivists employed a singular weapon: the principle of verification. Its content was: the meaning of a sentence is its method of verification. The principle was intended to establish a clear criterion on the basis of which one could accept a sentence as meaningful or reject it as nonsense.

In the case of mathematical statements, the method of verification was the construction of a proof. In the case of empirical statements, even as outlandish as 'On the back of the moon a Japanese magnolia is coming into leaf', the

principle leads us to conclude that this is a meaningful sentence, because we are able to stipulate a method of how we could verify the statement. But in the case of a statement like 'Being and Nothing are ultimately the same', a phrase the logical positivists interpreted as a metaphysical one, there is no way of verifying it. It, therefore, ought to be rejected as nonsense.

Straightforward as the principle might seem, it encountered an insurmountable objection. What if we apply the principle to its own content? How can we verify the principle of verification? It turns out that we cannot, and it must therefore be rejected as an article of faith. In philosopher's jargon: the principle is self-referentially incoherent. Logical positivism rested on a metaphysical assumption.

## 6. THE PHILOSOPHY OF LOGICAL ATOMISM

Logical positivists have often been accused of defending a shallow form of scientism. The book they hailed as their source of inspiration most certainly was not. The *Tractatus Logico-Philosophicus* by Ludwig Wittgenstein is one of the most dense, impenetrable and intriguing works of philosophy ever written. The logical positivists interpreted the following proposition from this work as, what they called, the above mentioned principle of verification: "To understand a sentence means knowing what has to be the case, if the proposition is true." (proposition 4.024). This proposition, however, ought to be read against the background of Wittgenstein's *picture theory of meaning*.

As we have seen, Russell argued in his *theory of descriptions* that the grammatical form of a sentence conceals its logical deep structure. This was a poignant criticism of Frege, but it leaves unanswered the question, to which Frege did propose an answer: what is meaning? Russell and Wittgenstein collaborated in the years preceding the First World War on a theory that should provide an answer to precisely that question. According to their *philosophy of logical atomism*, we should further pursue the analysis of the logical structure underneath the grammatical surface structure of a sentence. Somewhere in that depth language 'hits' reality. The logical atoms out of which a meaningful sentence is composed have a one to one relationship with the logical structure of the state of affairs it describes. This brought Wittgenstein to the view that, since ordinary language sentences are meaningful, such a sentence must be a logical picture of the state of affairs it describes. But what about the statements of logic, that is, statements that tell us something *about* the relationship between sentences? Are they too pictures of states of affairs? And what about ethical statements or other philosophical assertions that do not seem to state a matter of fact? Do they all represent states of affairs?

To account for the meaningfulness of logic Wittgenstein appeals to a distinction between *saying* and *showing*. Sentences can only be meaningful if they say something about the world. Logical sentences do not say something about the world. In acknowledging the truth of a logical tautology, we *show* that we have understood the meaning of the logical symbols, but we have not *said* anything about the world. The distinction can also be applied to ethics. An ethical statement doesn't say something about the world. Ethics is a matter beyond the limits of language: it has to be *shown*. You do not say that you are a good human being, you *show* that in your actions.

A question that is left unanswered in the *Tractatus* is what these mysterious logical atoms are. According to Russell, they are the simplest pieces of information that we acquire in sense perception. As he calls them: they are 'sense data', to which we refer with logical proper names like 'this' and 'that'. Sense data are the deliverances of the senses to the mind. It is a notion that Russell picked up from psychologists, true to his belief that philosophy is part of the unity of science.

## 7. MEANING IS USE

**A**fter the Second World War, A.J. Ayer, following Russell, employed the scientific notion of sense data<sup>8</sup> to give a foundation for our knowledge about the external world. According to Ayer this foundation is based on the deliverances of the senses, sense data, whether these represent reality truthfully in veridical perceptions or not, as in is the case in hallucinations. Ayer sees it as the task of philosophers to construct a coherent and consistent theory about the external world. First, philosophers should develop and employ a 'sense datum language' in terms of which we describe our perceptions: yellow-here, feeling so-and-so. Second, philosophers should translate this so called 'sense datum language' into ordinary 'object' language ('banana') and then examine which sentences of this language fit into a coherent theory about the world and which cannot and are presumably reports of illusions.

The first problem is the simple question of what such a sense-datum language amounts to. It has turned out that it is nearly impossible to design a language with words that only refer to sense data, like green-here, bitter-here. But let us suppose that it were possible to do so. What are then the requirements such a sense datum language would have to meet? First and foremost, since we

---

8. Sense data are supposed to be the mind-dependent objects that we are directly aware of in perception. For instance, if we perceive a banana in normal conditions, we are aware of a mental image that is yellow and has a peculiar curved shape. This mental image is composed out of 'sense data'.

cannot share our sense data, it needs to be an *essentially* private language, in the sense that only the person whose language it is can understand it. In his second major book, *Philosophical Investigations*, Wittgenstein argued that such a private language is impossible. Let us suppose, first, that it is possible to give meaning to words for private experiences. Whenever experience Y occurs, we say 'Oeph' in our private language. In ordinary life and language, we occasionally make mistakes. We think we see a green lemon, but it is a lime. After having examined the fruit we correct our mistake. This calibration on objects in reality is an essential element in speaking a language. Is such a correction procedure possible in a private language? Wittgenstein claims that it is not: "[...] I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'.

But there is an even more fundamental problem that Wittgenstein exposes at the beginning of the *Philosophical Investigations*. We assumed that we could give in our private language names to private experiences, supposedly in ostensive definitions, i.e. definitions of a word by pointing to the object or property that is the meaning of the word, for example 'This is yellow'. But the very idea that putting a name to an object (in general, not just to private experiences) as a kind of label to secure the tie between language and reality is a mistake. That there is more to meaning than being a label for an object (what we have called 'the naïve theory of meaning') Wittgenstein shows with a description of a particular situation in which certain words are used in a particular way (which Wittgenstein calls 'a language game'). Imagine a builder and an assistant. The builder shouts 'Hammer!'. Suppose all there is to meaning and understanding is a connection between the name and the object it stands for. In that case the assistant only has to establish a connection in his mind between word and the object hammer; he only has to know of what object the word hammer is a label. But, of course, the boss demands more, he wants his assistant to *bring* him the hammer, so he can apply this tool. By yelling 'Hammer!' he wants his assistant to *do* something.

Wittgenstein condensed his view in the slogan: "the meaning of a word is its use in the language." Wittgenstein thus seemed to pursue a middle course. On the one hand he rejected the idea that words could be tied to objects in reality via ostensive definitions; on the other hand, he rejected Locke's idea that words refer primarily to mental ideas in our minds. This raises the question where Wittgenstein stands in the realism versus idealism debate. Many interpreters attribute to Wittgenstein a version of *linguistic idealism*. What they mean by that can be elucidated with an example. Two people walk along the beach. Suppose one of them knows the meaning of the word wave, the other does not. "See how many waves are approaching the beach," the person who knows the meaning of the word wave exclaims. The other person only perceives a mass of dark, grey water. Knowing the meaning of the word wave enables someone to perceive



structure in an amorphous lump of grey water. Linguistic idealism is the position that understanding a language provides one with a capacity to impose particular structures on reality as we perceive it. In terms of the Fregean theory of meaning, described in this chapter: we first have to grasp the sense of a word, in order to be able to single out the object the word refers to.

The question now arises of how we get from the sense to the object referred to. A simple answer to that question is that the sense is a description of that object. However, we have learnt from Russell's criticism of Frege that descriptions cannot be equated with proper names. Boris Johnson will always be called 'Boris Johnson', but he will not always be the present prime minister of the UK. So that description is, at the time of writing, true of Johnson, but false at a date in the future.

This has led Saul Kripke to call proper names *rigid designators*. Kripke was a brilliant logician, who already as a student in high school laid the foundations for a formal system in which arguments that employ the notions of necessity and possibility could be examined. Rigid designators are expressions that refer to the same thing in every possible situation. 'Johnson' is therefore, a rigid designator, but 'the present prime minister of the UK' is not. Kripke, and his colleague, Hilary Putnam, claim that words for natural kinds such as water are also rigid designators. At some point in history an initial 'baptism' has taken place when someone named an object by pointing to it, e.g. "This is an elm". After that initial baptism ceremony later uses of the word 'elm' can all be brought back to this initial name-giving: there is a causal link through history that connects present people using the word 'elm' to that very same object that has been baptised with the natural kind term 'elm'. So, if someone now refers to an object by saying 'That is an elm', it has to belong to the very same natural kind as the object in the initial baptism ceremony, otherwise the word has not been used correctly.

This so-called 'causal theory of reference' brings us back to the view that names are labels we attach to objects that are already out there, waiting for their discovery. Kripke and Putnam, near the end of the 20th century, have reversed the linguistic turn. Their philosophy of language presupposes an outspoken form of realism: there is a ready-made world out there, consisting of discrete objects to which we assign words as labels.

## 8. CONCLUSION

The 'causal theory of reference' became the dominant theory at the end of the 20th century. The course of the history of philosophy of language thus was a full circle, because this meant a return to the idea that the meaning of a noun is the object it stands for, what we have called 'the naïve theory of meaning'. What have we learned along the way? We cannot but conclude that the philosophy of language cannot carry the weight it was supposed to do. The fundamental ontological question of whether realism or idealism about the external world is true cannot be answered within the philosophy of language. Nor can an analysis of thinking be complete, if it only analyses thoughts that can be put into words. So, the fundamental tenets of the philosophy of language with which we started this chapter have to be rejected. But what has been achieved by focussing on language is that philosophy has become much more precise, technical in certain respects, and more aware of its own methodology. Philosophy of language has produced indispensable tools, like Frege's distinction between sense and reference, and awareness of the distinction between talking about words and talking about things, for doing philosophy in many of its sub-disciplines, like meta-ethics, metaphysics and the philosophy of mind. In general, it has made philosophers even more sensitive to the important but sometimes elusive role of language in the way in which we phrase our philosophical ideas, problem and theories. As an offshoot: although the philosophy of language in the broad sense might no longer be at the centre stage of philosophy, contemporary philosophy of language (in the narrow sense as 'philosophy of linguistics') is flourishing, as a branch of science independent from philosophy.<sup>9</sup>

---

9. I am much obliged to dr. Anne Sophie Meincke for many helpful comments on an earlier draft of this chapter.

## BIBLIOGRAPHY

- Austin, J. (1961). A plea for excuses. In J. O. Urmson & G. J. Warnock (Eds.), *Philosophical papers* (pp. 129–130). Oxford University Press.  
<https://doi.org/10.1093/019283021x.003.0008> (Original work published 1956)
- Austin, J. (1963). *How to do things with words* (J. O. Urmson, Ed.). Oxford University Press. (Original work published 1962)
- Ayer, A. (1940). *The foundations of empirical knowledge*. MacMillan & Co.
- Ayer, A. (1980). *Language, truth, and knowledge*. Penguin Books Ltd. (Original work published 1936)
- Dummett, M. A. E. (1978). Frege's philosophy. In *Truth and other enigmas*. Duckworth.
- Dummett, M. A. E. (1991). *Frege. Philosophy of Mathematics*. Duckworth.
- Dummett, M. A. E. (1993). *Origins of Analytical Philosophy*. Duckworth.
- Frege, G. (1952). On sense and reference. In D. Byrne & M. Kolbel (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 49–55).
- Kripke, S. (1980). *Naming and necessity*. Basil Blackwell.
- Locke, J. (1985). *An essay concerning human understanding*. Oxford University Press. (Original work published 1690)
- Putnam, H. (1975). The Meaning of Meaning. *Minnesota Studies in the Philosophy of Science*, 7, 131–193. [https://doi.org/10.1007/978-94-009-5496-0\\_3](https://doi.org/10.1007/978-94-009-5496-0_3)
- Rorty, R. (1991). Wittgenstein, Heidegger, and the reification of language. In *The cambridge companion to heidegger* (pp. 337–357). Cambridge University Press.  
<https://doi.org/10.1017/ccol0521385709.014>
- Russell, B. (1984). On denoting. In R. C. Marsh (Ed.), *Logic and knowledge* (pp. 41–56). George Allen & Unwin. (Original work published 1905)
- Ryle, G. (1949). Review of 'Meaning and necessity'. *Philosophy*, 24(88), 69–76.  
<https://doi.org/10.1017/s0031819100006781>
- Searle, J. (1976). *Speech acts*. Cambridge University Press. (Original work published 1969)
- Strawson, P. (1971). On Referring. In *Logico Linguistic Papers* (pp. 1–27). Methuen. (Original work published 1950)
- Wittgenstein, L. (2009). *Philosophical investigations* (G. E. M. Anscombe, J. Schulte, & P. M. S. Hacker, Trans.). Wiley-Blackwell. (Original work published 1953)
- Wittgenstein, Ludwig. (1922). *Tractatus Logico-Philosophicus* (C. Ogden & F. Ramsey, Trans.). Kegan Paul. (Original work published 1918)

LUDWIG WITTGENSTEIN



# LUDWIG WITTGENSTEIN

## PHILOSOPHICAL INVESTIGATIONS

---

The *Philosophical Investigations* (1953) by Ludwig Wittgenstein consists of aphorisms, most of them short, often in the form of a dialogue. However, in the text an intricate structure can be discerned, which allows for a division of the book in 'conventional' chapters. The best way towards understanding the book is by first reading the aphorisms on the nature of philosophy (§§ 89-133). Wittgenstein argues that philosophy is not a theoretical enterprise, but a therapeutic activity that aims to dissolve philosophical problems. The therapy that Wittgenstein advocates is to look carefully how words are being used in ordinary language. By describing minutely this everyday use problems disappear. At the beginning of his book Wittgenstein identifies one crucial mistake: the idea that the meaning of words is the object they stand for. In the philosophy of mind this gives rise to the mind-body problem: if the word body is meaningful, because it stands for an object, then the meaningful word mind also has to stand for an object. And then we end up with the problem of how these two objects are related. In the philosophy of mathematics it leads to the view that numbers are objects. If the phrase 'the number three' is meaningful, then there has to be an object number three. Wittgenstein opposes this idea; he emphasizes that words are *used* in daily life to *do* things. Of course, not any use is meaningful; use has to be *rule*-governed. He, therefore, devotes central passages to what it means to follow a rule. These considerations evolve into arguments against the possibility of a private language, which in turn lead Wittgenstein to reflect on thought, imagination, intentionality, and other topics in the philosophy of mind. The book is an organic whole that demands and repays careful study.

---

# ARTWORK ACKNOWLEDGEMENTS

The images within this book are drawings made by the Tilburg based artist S. Lloyd Trumpstein, a.k.a. Stephan de Weert. The drawings are tributes of iconic portraits of philosophers. Most of the originals are part of the public domain, but some are not. We did our best to find the copyright holders. If you think your material is used without permission, we kindly invite you to get in touch with us: [info@openpresstiu.org](mailto:info@openpresstiu.org).

1. Fred Stein, Hannah Arendt, 1949. With permission of the Fred Stein Archive/Getty Images.
2. [Photographer Unknown], Virginia Held, [19xx]. With permission of Virginia Held.
3. Claude Truong-Ngoc, H el ene Cixous, 2011. [Wikimedia Commons/Public Domain](#).
4. G. Edelinck, Blaise Pascal, after Monogrammist LVQ, 1666 – 1707. Public Domain.
5. Jonas Suyderhoef, Portrait of Ren e Descartes, after Frans Hals, 1650 – 1675. Public Domain.
6. Irving Penn, Simone de Beauvoir, 1957. With permission of the Irving Penn Foundation.
7. Sally Ryan, Martha C. Nussbaum, 2010. [Wikimedia Commons/Public Domain](#).
8. [Photographer Unknown], Edmund Husserl, [19xx]. Public Domain.
9. Raphael, The School of Athens [detail of Plato], 1509-1511. Public Domain.
10. Bill Pierce, Thomas Kuhn (1922–1996), [Wikimedia Commons/Public Domain](#).
11. [Photographer Unknown], Rose Rand, 1949. [Wikimedia Commons/Public Domain](#).
12. [Artist Unknown], John Stuart Mill, circa 1870. Public Domain.
13. [Photographer unknown], Willard van Orman Quine, 1975. [Wikimedia Commons/Public Domain](#).
14. Willy Pragher, Martin Heidegger, 1960. [Wikimedia Commons/Public Domain](#).
15. Moritz N ahr, Ludwig Wittgenstein, 1930's. [Wikimedia Commons/Public Domain](#).