Science and Imagination

Sam Rijken

Science and Imagination

Sam Rijken

This work was supported by the Erasmus School of Philosophy.

Title page designed by D.C. van Houten - www.dcvanhouten.com.
Printed by Print Service Ede - www.proefschriftprinten.nl.

ISBN: 978-90-834134-3-3

Science and Imagination

Wetenschap en voorstelling

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Breedenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Friday 3 May 2024 at 10:30 hrs

by

Sam Michael Rijken

born in Ede, The Netherlands.

Doctoral Committee:

Promotors:

Prof.dr. F.A. Muller

Dr. T.K.A.M. De Mey

Other members:

Prof.dr. A.J.M. Peijnenburg

Dr. M.T. Stuart

Dr. S. Wintein

Contents

List of Figures

Preface

Prelude

Hi, my name is Sam. Welcome to my thesis: *Science and Imagination*. When I began my PhD research in February 2019, I originally set out to study and understand scientific thought experiments. In particular, I aimed to formulate a clear answer to the following research questions:

> What are scientific thought experiments and what, and how, can we learn about the natural world by performing them?

This much is clear: thought experiments are a tool of the imagination. But what in tarnation is *imagination*? I quickly found out that nobody really knows. Unacceptable. And so the focus of my research shifted from scientific thought experiments in particular to imagination in general. I aimed to understand scientific thought experiments and their epistemic value, but I first had to understand imagination and *its* epistemic value:

> What is imagination and what, and how, can we learn about the natural world by using it?

My research proceeded thusly.

Before I dive into the details of this Thesis, I wish to illustrate my fascination for imagination by walking you through my favorite poem about imagination: Jorge Louis Borges' *The Other Tiger*.[1] Admittedly, I interpret this poem as being about imagination, but I am not entirely certain that Borges intended it as such — *I imagine it so*, which works for me.

In *The Other Tiger*, Borges beautifully describes how he becomes aware that there are *three* distinct 'types' of tiger. First and foremost, Borges begins, there is the *real tiger* — the living, flesh-and-blood, ferocious and awe-inspiring organism of near-infinite biological complexity which roams the Earth:

> A tiger comes to mind. The twilight here
> Exalts the vast and busy Library
> And seems to set the bookshelves back in gloom;
> Innocent, ruthless, bloodstained, sleek
> It wanders through its forest and its day
> Printing a track along the muddy banks
> Of sluggish streams whose names it does not know
> (In its world there are no names or past
> Or time to come, only the vivid now)
> And makes its way across wild distances
> Sniffing the braided labyrinth of smells
> And in the wind picking the smell of dawn
> And tantalizing scent of grazing deer;
> Among the bamboo's slanting stripes I glimpse
> The tiger's stripes and sense the bony frame
> Under the splendid, quivering cover of skin.
> Curving oceans and the planet's wastes keep us
> Apart in vain; from here in a house far off
> In South America I dream of you,
> Track you, O tiger of the Ganges' banks.

[1] Obtained from https://www.blueridgejournal.com/poems/jlb-tigr.htm, last accessed 15 September 2023.

Upon writing the passage above, Borges realises that, when he tries to *describe* a real tiger using words, he will never do full justice to the 'real thing'. When he tries to describe a tiger, when he tries to *capture it* on paper, he will only produce a string of symbols, a mere *description* or representation of a tiger. He may produce a tiger of symbols, but he will never get the real deal. This rather disappoints Borges — understandably so, in my view, as I have experienced the same disappointment many times[2] — as it seems that there is no way to bridge the gap between the 'real thing' and our necessarily limited symbolic description thereof. Thus, secondly, directly opposed to the real tiger, Borges finds the *tiger of symbols*:

> It strikes me now as evening fills my soul
> That the tiger addressed in my poem
> Is a shadowy beast, a tiger of symbols
> And scraps picked up at random out of books,
> A string of labored tropes that have no life,
> And not the fated tiger, the deadly jewel
> That under sun or stars or changing moon
> Goes on in Bengal or Sumatra fulfilling
> Its rounds of love and indolence and death.
> To the tiger of symbols I hold opposed
> The one that's real, the one whose blood runs hot
> As it cuts down a herd of buffaloes,
> And that today, this August third, nineteen
> Fifty-nine, throws its shadow on the grass;
> But by the act of giving it a name,
> By trying to fix the limits of its world,
> It becomes a fiction not a living beast,
> Not a tiger out roaming the wilds of earth.

[2] 23-year old me said defiantly (and meaninglessly): "*It's all language games, man.*"

But this is not the end of the story. Borges now insists that there is *another* tiger alongside the two tigers just discussed. A third one, a tiger that does not stand opposed to either the real tiger or the tiger of symbols, *the tiger of our imagination*:

> We'll hunt for a third tiger now, but like
> The others this one too will be a form
> Of what I dream, a structure of words, and not
> The flesh and one tiger that beyond all myths
> Paces the earth. I know these things quite well,
> Yet nonetheless some force keeps driving me
> In this vague, unreasonable, and ancient quest,
> And I go on pursuing through the hours
> Another tiger, the beast not found in verse.

Borges admits that this Other tiger is more closely related to the tiger of symbols than it is to real tigers, but nonetheless insists that the Other tiger and the tiger of symbols are distinct. Beyond this, Borges remains silent on what this Other tiger amounts to, but I would say something along the following lines: whereas the tiger of symbols consists merely of symbols, the tiger of my imagination consists not only of symbols but also of *my personal experiences, memories, beliefs, associations, fears and desires* and so on. The tiger of my imagination is, indeed, the tiger of *my* imagination. *I am part of the Other Tiger.* That is what distinguishes the Other Tiger, the tiger of my imagination, from the real tiger and the tiger of symbols.

Before I move on, I wish to note that, for me, this poem highlights the irony with *analyzing* imagination, certainly with studying it using the tools of analytic philosophy: when I set out to *describe* and *analyse* imagination, my result will always be a 'mere' string of symbols, it will never capture 'imagination itself'. I will always come short. And yet, nonetheless some force kept driving me in that vague, unreasonable and ancient quest, and I go on pursuing through the hours: *imagination.*

Acknowledgements

This Thesis would not have existed without the endless support of the people around me. I next acknowledge those who helped me the most.

First and foremost, I wish to thank my PhD supervisor prof.dr. F.A. Muller. Fred: your clarity of thought, admirable work-ethic, and never-ending battle against ambiguity in philosophy has shaped me into the philosopher that I am today. I cannot *conceive* (Section 2.5.3) what my Thesis would have looked like without your excellent supervision.

Next, I thank dr. Noelia Iranzo Ribera and soon-dr. Nick Wiggershaus. You are both good friends and I was incredibly fortunate that the topics of your PhD's were similar to mine. I thoroughly enjoyed exploring with you — both in informal and academic settings — the ramifications of introducing the concept of *fiction* in philosophy of science.

In this same manner, I thank all academics with whom I interacted along the way. I especially thank prof.dr. Frigg, dr. De Mey and dr. Stuart, who provided incredibly helpful comments on early drafts of the Chapters that make up this Thesis. I also thank the members of my doctoral committee, prof.dr. Peijnenburg, dr. Stuart and dr. Wintein, for their thorough reading of this thesis and helpful feedback. I also thank soon-dr. Ruward Mulder, with whom I co-organised a reading group on thought experiments, which was attended by several 'big shots' of the field.

I also thank Daan van Houten, good friend and artist, who designed the cover of this Thesis after I gave him the intentionally-impossible task: "draw me Borges' Other Tiger".

I thank Kenny, who died too young but changed my life nonetheless. Kenny: the way you lived life relentlessly will continue to amaze and inspire me for the rest of my life. I hope that I can make you proud.

Finally, I thank my parents Harry and Maud, my sister Kim, my girlfriend Esme, and my friends, many of whom I have known since kindergarten. I know how lucky I am to have you. I think about it every day. Thank you.

Chapter 1

Introduction

1.1 *Science and Imagination*

1.1.1 Research questions

This Thesis consists of three main Chapters: *Explicating Imagination* (Chapter 2), *Knowledge Through Imagination* (Chapter 3), and *Scientific Thought Experiments* (Chapter 4).

In the first main Chapter, *Explicating Imagination*, I deal with the following three research questions:

- ➢ How can we explicate the mental state of imagination?
- ➢ What are core characteristics of the mental state of imagination?
- ➢ How does imagination relate to similar mental states, notably to perception, belief, visualisation, supposition and memory?

In the second main Chapter, *Knowledge Through Imagination*, I deal with only one research question:

- ➢ Is imagination a source of knowledge of the natural world?

Finally, in the third and last main Chapter, *Scientific Thought Experiments*, I deal with two research questions:

- ➢ What are scientific thought experiments (STEs)?
- ➢ What, and how, can we learn by performing STEs?

1.1.2 Methodology

This is an analytic-philosophical Thesis. Much of the analysis provided in this Thesis comes forth from *armchair inquiry*: the method of my research is mainly *conceptual analysis* on the basis of literature review. No empirical research has been performed as part of this Thesis — but, of course, as a philosopher *of science*, I always make sure that my results do not *conflict* with, but rather *are informed by* and *complement*, current scientific knowledge and understanding of the relevant topics.

I discuss many different topics in this Thesis, pertaining to e.g. imagination, perception, memory, knowledge, fiction and models. Entire monographs have been written about each and every one of these topics. I cannot do justice to all existing literature in this single Thesis. But that is not my aim. My aim, rather, is to bring *much* of the literature together into a single, *coherent whole*. It is often lamented that the concept of imagination *resists* unambiguous philosophical analysis. I disagree: in this Thesis, I show that it is possible to analyse imagination unambiguously. My analysis will presumably raise as many questions as it attempts to answer. My aim is achieved if those questions can now be asked *clearly*.

1.1.3 Overview of Chapters

Chapter 2: Explicating Imagination

In this Chapter, I propose explications for the concept of *imagination* and many of its closely-related concepts. I begin by distinguishing imagination from *perception*, *optical illusions*, and *hallucination*. I then distinguish two *types* of imagination: *proposition*-imagination and *action*-imagination. I then first provide an explication for proposition-imagination, and I discuss how this explication holds in light of — and sheds a new light on —

eight 'core characteristics' that are often associated with imagination in the literature. Using this explication, I then explicate the concepts of *supposition*, *counterfactual thought*, *conceiving*, *visualisation* and *picturing* as types of proposition-imagination. I then turn to explicating the second type of imagination: action-imagination. Using this explication of action-imagination, I revisit what it means to visualise and picture actions, and I relate imagination to *memory*. Finally, I comment on characteristic aspects of imagination in practice, and I provide some brief but necessary notes on the cognitive science of imagination.[3]

Chapter 3: Knowledge Through Imagination

In this Chapter, I discuss how imagination can function as a source of knowledge of the natural world. I begin by explicating, in contrast to 'ordinary' perception, the concept of *quasi-perception*, i.e. the 'perception-like' mental state that we have when we *imagine* perceptions or vividly *remember* the past. I provide a two-step framework for how we obtain novel *beliefs* about the natural world on the basis of quasi-perceptions, which I call *quasi-perceptual beliefs*. I then discuss at length how quasi-perceptual beliefs are epistemically *justified*. I then discuss how *imagination* can be responsible for this justification. Finally, I distinguish and discuss several senses of the term "source of knowledge". I conclude that (i) imagination is not a so-called *basic* source of knowledge, (ii) imagination is certainly what I call a *crucial* source of knowledge, and (iii) imagination is even what I call a source of *otherwise-inaccessible* knowledge.[4]

Chapter 4: Scientific Thought Experiments

In this Chapter, I discuss what scientific thought experiments (STEs) are and what, and how, we learn by performing them. I introduce several example STEs, each of which serve to illustrate important characteristics

[3] This Chapter is partially based on a draft paper, co-authored with F.A. Muller. [4] A small part of this Chapter (Sections 3.4.2–3.4.6) is based on a draft paper, co-authored with F.A. Muller.

of STEs. I then elaborate on the two research questions mentioned-above, and I discuss two long-standing accounts of STEs — the argument view and the mental-modeling view — indicating their strengths and weaknesses. I then introduce the theory of fiction from Walton (1990) and discuss two recently proposed accounts of STEs that are explicitly built on this theory of fiction. To improve on these recent proposals, I then introduce the *fiction view of models*, which I use to formulate a full-fledged account of STEs: the *fiction view of scientific thought experiments*.[5]

1.2 A brief history of imagination

To kick off this Thesis, I provide a very brief history of imagination. Our understanding of imagination and related concepts has undergone several important transformations since the inception of Western philosophy over two-and-a-half millennia ago. I next indicate key developments that are important to keep in mind when reading this Thesis.

1.2.1 Imagination from Aristotle to Kant

Ever since the inception of Western philosophy in and around Ancient Greece, imagination has played a crucial role in the philosophical method. Allegories, metaphors, thought experiments and other forms of arguments that strongly appeal to the imagination were regularly employed by pre-Socratic philosophers — think, for example, of Zeno's paradoxes. This method culminated in the work of Plato: the undisputed king of allegories and thought experiments.

However, whereas Plato *appealed* to imagination constantly throughout his works, he did not extensively *analyse* imagination, at least not explicitly (Bundy, 1922; Hart, 1965; Wedgwood, 1977). As such, Plato's

[5] This Chapter is based on a paper which was submitted to (and rightly rejected by) *BJPS* in 2019. To improve the quality of this Chapter, I have used the anonymous referee reports from that submission and private feedback from notably Roman Frigg, Mike Stuart, Tim De Mey, the participants of the *Working Models* reading group, amongst many others.

conception of imagination — if there was one — is largely lost to history. Instead, it was the conception and explicit discussion of imagination by Aristotle "which came down through the Middle Ages as the accepted tradition" (Bundy, 1922, p.262) and which remains relevant to this day.[6]

Aristotle discussed the concept of imagination explicitly in *On The Soul* (*De Anima*), straight after discussing the nature of the five senses and right before discussing the nature of thought. (This sandwiched position of imagination in between sense-perception and thought is no coincidence). Aristotle put forward a conception of imagination that has remained the standard for over two millennia (Book III, part 3):[7]

> imagination is that in virtue of which an image arises for us

For Aristotle, imagination is a mental faculty that underlies and assists our memories, dreams and thoughts by providing them with mental imagery — but not our perceptions, as they provide their own imagery. In other words: *imagined* mental imagery is imagery-in-absence-of-perception.

Beyond this, Aristotle remains relatively brief on the matter, spending most of his time trying to distinguish imagination from, on the one hand, ordinary sense perception, and, on the other hand, belief (or "thought"). After concluding that imagination is indeed distinct from ordinary sense perception and belief (or any combination thereof; c.f. Shields (2020)), Aristotle submitted defeat in further clarifying the concept:

> About imagination, what it is and why it exists, let so much suffice.

Let us now make a big jump through history and land at the next philosopher who importantly transformed our understanding of imagina-

[6] Bundy (1922, p.362) explains: "The reasons for this comparative neglect of Plato are not far to seek. The directness of Aristotle's method in comparison with the subtle art of the Dialogues rendered the views of the former much easier of comprehension. Much of the suggestiveness of the Platonic conception [of imagination], one fears, has been lost through lack of sympathy with the artistic purposes of the philosopher-poet." [7] I use the English translation of *De Anima* from Aristotle (2022), accessed 6 July 2023.

tion:[8] Descartes. For Descartes, too, imagination was intimately tied to mental imagery (*Meditations*, VI 72):[9]

> [W]hen I imagine a triangle, not only do I understand it to be a shape enclosed by three lines, but at the same time, with the eye of the mind, I contemplate the three lines as present, and this is what I call imagining.

Importantly, Descartes presented not only a conceptual analysis of imagination (by discussing its relation with perception and reason, or "pure understanding", like Aristotle did) but also an *epistemological* analysis of it. True to his sceptic self, Descartes expressed great scepticism about the ability of imagination to produce "certain and evident knowledge of the truth" (*ibid.*, VI 10). In fact, "like other rationalists, Descartes dismisses imagination as the wrong kind of faculty to produce the secure knowledge that he seeks" (Kind and Kung, 2016, p.6). After having presented an argument for the existence of external objects that directly appeals to imagination, Descartes noted (*Meditations*, VI 73):

> I therefore conclude with great probability that the body exists. But this is only a probability, and although I am investigating the whole matter with great care, I do not yet see that, from this distinct idea of bodily nature that I find in my imagination, any argument can be derived that will lead necessarily to the conclusion that some body exists.

So, for Descartes, imagination can give (in modern terminology) *credence* to certain beliefs, but it can never guarantee their truth, and, hence, it cannot truly ground *certain* knowledge of nature. The battle between imagination and knowledge had begun, and it would only grow fiercer.

[8] I note that Ibn Sînâ (Avicenna, 980–1037AD) had an exceptionally sophisticated view of imagination, the likes of which we would not see until Hume and Kant. It has been argued that Ibn Sînâ's view of imagination had limited influence on later philosophers, c.f. Portelli (1979); Black (1993); Bäck (2005); Yaldir (2009). It can also be argued, however, that Avicenna's (proto) 'cognitive psychology' greatly influenced e.g. Fodor's (1983) concept of "modularity of mind" and the conception of "cognitive faculties" as a whole; e.g. Perler (2015); Silva (2020). (I thank Tim de Mey for this last remark.)

[9] I use the English translation of the *Meditations* from Descartes (2008).

Let me illustrate this by briefly discussing two other rationalists' views on imagination: Spinoza's and Baumgartner's views.

First: Spinoza. Spinoza greatly amplified Descartes' epistemological scepticism about imagination by arguing not only that imagination cannot ground certain knowledge but, rather, that imagination *is* "the source of error about the very nature of things" (Kind and Kung, 2016, p.8):

> And since those who do not understand the nature of things, but only imagine things, make no affirmative judgments about things themselves and mistake their imagination for intellect, they are firmly convinced that there is order in things, ignorant as they are of things and of their own nature. [...] And since those things we can readily picture we find pleasing compared with other things, men prefer order to confusion, as though order were something in Nature other than what is relative to our imagination.

There are two important things to note about this passage. Firstly, Spinoza regards imagination not only as the source of imagery-in-absence-of-perception, but also as the prime source of *false* beliefs about Nature. This idea — opposing imagination to *truth* — persists to this day. Secondly, Spinoza here regards imagination as the source of perceived "order in things", i.e. as the source of perceived structure in Nature, which in reality lacks such structure. This is an important idea that would later be immortalised in the form of Kant's *schemata* (discussed below).

Second: A.G. Baumgarten. The Enlightenment rationalist philosopher and founding father of Aesthetics, A.G. Baumgarten, expressed epistemic scepticism about imagination from a different angle than Spinoza did, but this angle was no less sharp. Baumgarten argued that imagination was only capable of a rather trivial re-presentation of memories. In doing so, he seemed to cast serious doubt over the ability of imagination to generate genuinely *new* ideas (*Metaphysik* (§414)):[10]

> And since my imaginations are representations of such things which used to be present (§210), they represent things which I have ex-

[10] Translation by Humber van Straalen, private communication.

> perienced [*empfunden*], but which are absent at the moment that I imagine them (§148). [...] As a consequence the power of imagination exclusively repeats representations and contains nothing, except that which has previously been in the senses.

But not all Enlightenment philosophers had such a sceptic view on imagination. Opposed to the Enlightenment rationalists stood the empiricists, who had very different conceptions of imagination than the rationalists did. The key figure to discuss here is, of course, David Hume.

Hume thoroughly transformed our understanding of imagination in several ways which remain relevant to this day. For Hume, too, imagination was responsible for the generation of mental imagery (in absence of perception). But for Hume, according to whom *all* thought has an imagistic format, imagination assumed deeply fundamental epistemic value: together with memory, imagination was essentially responsible for *giving shape to, and making possible, all thought.*[11] Diving deeply into Hume's multifaceted conception of imagination and its *sine qua non*-role for thought would be the topic of an entire Thesis; c.f. Streminger (1980); Traiger (2008); Wilbanks (2012); Cottrell (2015); Dorsch (2016a). This I cannot and will not do. Instead, I limit myself here to noting three contributions that Hume made to our understanding of imagination that are directly relevant for the purpose of Thesis.

Firstly, as I indicated above, Hume tied our faculties of imagination and memory close together. For Hume, both imagination and memory had the role of providing our thought with its imagistic form and content. The big question that presented itself, then, is what the *difference* is between memory and imagination. Hume's answer is famous[12] (*Treatise*, §1.1.3)[13]:

> When we remember any past event, the idea flows in upon the mind

[11] As Kind and Kung (2016, p.8) note: "The importance of imagination in Hume's cognitive psychology, especially in contrast to rationalists like Descartes, would be hard to overstate. According to Hume's Copy Principle, all mental contents are in some sense imagistic, and as such the imagination lies at the foundation of his cognitive psychology."

[12] Albeit no longer accepted in contemporary literature, as will transpire throughout this thesis, notably Chapter 3; c.f. Urmson (1967); Huemer (2001).

[13] I use Hume (1896) for quotes from the *Treatise of Human Understanding*.

> in a forcible manner; whereas in the imagination the perception is faint and languid, and cannot with difficulty be preserv'd by the mind steddy and uniform for any considerable time.

For Hume, memories are *forceful and vivid*, whereas imaginings are *faint and languid*. This suggests that Hume does not consider there to be a *sharp* distinction between memory and imagination. In modern terms, we would describe this as a distinction on a *phenomenological* level, rather than at the level of topic, content or format of "ideas" (i.e. mental states).

Secondly, and most importantly, Hume explicitly connected imagination to *possibility*, which idea is immortalised in his famous statement that (*Treatise*, §1.2.2, original italics):

> 'Tis an establish'd maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence*, or in other words, *that nothing we imagine is absolutely impossible.*

The link between the concepts of imagination and possibility that Hume forged proved convincing and remains relevant to this day.[14] Two important consequences are (i) that any conception of imagination *must* explicitly connect it to possibility, and (ii) from this idea sprung forth an entire branch of literature that concerns the question whether imagination is a source of *modal* knowledge, i.e. knowledge of (im)possibilities.[15]

Thirdly, and closely related to the previous point, Hume put forward a nuanced account of the *powers* and *limits* of imagination. To begin: imagination is *limited* by the resources it has at its disposal, which is the sum-total of our past (sense) experience. These are the "ideas" available for our imagination to work with. But imagination is *unlimited* in its power to *re-combine* these past experiences in all trivial and non-trivial ways,

[14] Although consensus is growing among contemporary authors that we can also imagine *impossibilities* (by regarding them *as* possible, White (1990) adds); c.f. (Berto, 2022, §5.1) and the references therein; and see Chapter 2 of this Thesis. [15] Hume argued that imagination *is* a direct source of modal knowledge (*Treatise*, §1.2.2): "We can form the idea of a golden mountain, and from thence conclude that such a mountain may actually exist. We can form no idea of a mountain without a valley, and therefore regard it as impossible." C.f. Tidman (1994).

wholly or in parts, continuous or fragmented, evident or surprising. Today, this account of the powers and limits of imagination is known as Hume's *recombination principle* — remember this phrase, I shall occasionally refer back to it throughout this Thesis.

The last influential author on imagination that I wish to discuss in this brief historical overview is Kant. For Kant, imagination arguably had an even more fundamental role than it did for Hume. To explain why this is the case, I must first introduce a few concepts from Kant's complex conceptual castle.

In his *Critique of Pure Reason*, Kant famously acknowledged two basic cognitive faculties: understanding (*Verstand*), i.e. the faculty responsible for conceptual thought, and "sensibility" (*Sinnlichkeit*), i.e. the faculty responsible for intuition, sense perception and mental imagery. Alongside these two cognitive faculties — or, more precisely, "*caused* by the action of the understanding on sensibility" (Hanna, 2022, §1.1) — Kant placed imagination (*Einbildungskraft*). Here, imagination is responsible not only for producing mental imagery, but also for the deeply fundamental task of generating the synthesizing '*schemata*' that, for Kant, make possible and give shape to all cognition. Kant writes (*Prolegomena*, §35):

> Synthesis in general, as we will later see, is an effect of the imagination alone, a blind but indispensable function of the soul without which we would have no cognition at all, but of which we are hardly ever conscious.

In the 'Kantian tradition', then, imagination can be understood as being a cognitive faculty that is inextricably interwoven with our faculties for thought and sensibility. The modern 'Kantian tradition' of investigating imagination proceeds along this line: it focuses mainly on the complicated *interaction* between the cognitive faculty of imagination and other cognitive faculties. Notably since Strawson (1970), for example, it is often argued that "most if not all perceptual experiences are infused with imagination" (Brown, 2018, p.133). This ascribes unique and deeply fundamental epistemic value to imagination. One should imagine Kant

nodding approvingly.

Imagination is now, in a sense, hidden away, buried deep down in the very foundations of our cognition: always inextricably connected with and often indistinguishable from everything else. And there imagination remained. Until quite recently, when, in late 20th-century analytic philosophy, imagination rapidly took center stage — this time, not as a cognitive faculty, but as a distinct type of *mental state*.[16]

1.2.2 Imagination in analytic philosophy

During the 20th century, a curious transformation took place in the way we construed, studied and understood imagination and other 'mental concepts' in philosophy. Due to Frege, Russell, Wittgenstein and the logical positivists, analytic philosophy had arisen. In analytic philosophy, sub-disciplines such as philosophy of mind, epistemology and many other branches of philosophy that are directly relevant for studying imagination, assumed radically different forms and proceeded via radically different methods than before.

Analytic philosophy emphasises philosophy of *language* (think about Borges' *tiger of symbols*) and focuses predominantly on the (formal) analysis of concepts. Analytic philosophy of mind focuses predominantly on analyzing *mental states* rather than cognitive faculties. 'Cognitive faculty' is somewhat of a woolly concept that is difficult to pin down (what is a cognitive faculty?; which faculties are there?; where is the faculty of imagination located?; what it its function?; how does it relate to other faculties?). Mental states are arguably more clearly delineated than cognitive faculties, conceptually speaking. A *mental state* — the 'state of mind' of a subject — is uniquely characterised by only a handful of characteristics,

[16] A terminological note: in English, the word "imagination" refers to both imagination as a cognitive faculty and imagination as a mental state, which is rather confusing. In Dutch, the difference between the *cognitive faculty* of imagination and a *mental state* of imagination is reflected in language: the former is translated as *verbeeldingskracht* or *voorstellingsvermogen*, whereas the latter is translated as *verbeelding* or *voorstelling*. Likewise for German: *Vorstellungskraft* or *Einbildungskraft* versus *Vorstellung* or *Einbildung*.

most notably: (i) the *content* of the mental state, (ii) the *intentional objects* of the mental state (what the content is 'about'), and (iii) the *attitude* that the subject adopts towards this content.

I illustrate the notion of a mental state by briefly discussing the mental states of belief and perception, both of which are closely related to the mental state of imagination in their own way, as we saw in the previous Section.

A mental state of belief (i) has *semantic content* (i.e. it has linguistic *meaning* and can often be evaluated as *true* or *false*), (ii) always has an intentional object (i.e. you have a belief *about something*), and (iii) the attitude that we adopt to the content of our beliefs it that we *think that it is true*. To provide some contrast: a mental state of e.g. *doubt* is the same as the mental state of belief in features (i) and (ii), but, in doubt, the attitude that we adopt towards this content (iii) is different: when we doubt something, we think that it is more likely *false* than true.

A mental state of *perception* is different from a mental state of belief in various ways. A mental state of perception (i) has *sensory content* (i.e. our sensory organs are relevant for mental states of perception, and these mental states are importantly accompanied by so-called *qualia*, which is the experience of 'what it is like' to perceive something), (ii) usually, but not always, has an intentional object, and (iii) the attitude that we adopt towards this content is that we think that it represents something that is *there, in front of us, like that, in the world*.

Of course, mental states of belief and perception are *connected* to each other, in the sense that perceptions give rise to (perceptual) beliefs, and beliefs influence what we perceive. Notwithstanding, the two are analysed as *distinct* mental states. It is the task for *epistemology*, then, to investigate how these distinct mental states interact in ways that give rise to *knowledge* and *understanding*.

So too is imagination in analytic philosophy analysed as a distinct type of mental state. Just like you can have a mental state of belief, perception, memory, desire, hope, pain or what have you, you can also have a

mental state of imagination. Kieran and McIver Lopes (2003, p.9) even asserted boldly that "it is an axiom of contemporary theories of imagining that states of imagination are mental states with propositional [i.e. semantic] contents." (In the next Chapter, I shall discuss how Kieran and McIver Lopes are almost, but *not quite*, correct.) This tradition investigates imagination not by placing it alongside other cognitive faculties such as perception and reason but rather alongside other *mental states*, notably belief, memory and mental states of perception.

Because mental states are characterised by (i) their *content*, (ii) their *intentional objects*, and (iii) the *attitude* that a subject adopts towards this content, the main question pertaining to imagination in this tradition, then, is: what uniquely characterises mental states of *imagination*, i.e. which types of (i) content, and (ii) intentional objects, and (iii) attitudes uniquely characterise mental states of imagination? As we shall see in Chapter 2, much of the debate here centers around the questions (a) whether or not imagination requires a mental *image*, i.e. whether or not a mental state of imagination necessarily has sensory content, and (b) in which way, and to what extent, imagination is a *voluntary* mental state, i.e. whether we are free to choose the content of our mental state of imagination ourselves, and whether the attitude that we adopt towards the content of a mental state of imagination is a *voluntary* attitude. (To provide some contrast: for mental states of perception and belief, typically neither the content (i) nor the attitude (iii) of these mental states is voluntary. This marks an important difference between imagination on the one hand, and perception and belief on the other.)

Once we have an answer to these questions, and once we understand more generally what uniquely characterises the mental state of imagination, we can begin to understand how imagination relates to *other* types of mental states. (Chapter 2 of this Thesis is directly devoted to providing an answer to these questions.) Then, once we have an answer to *those* questions, we can start to do *epistemology* of imagination — that is, we can begin to understand how the mental state of imagination in-

teracts with other mental states in ways that give rise to knowledge and understanding. (Chapter 3 of this Thesis is devoted to this question.)

In line with contemporary analytic tradition, the topic of my analysis is the mental state of imagination, not the 'deep' *cognitive faculty* of imagination that was the topic of investigation in the 'Kantian tradition' that I described in the previous Section. It should be clear, therefore, that the topic of my investigation — imagination as a mental state — does not concern *every* aspect of the concept of imagination. Such is the way of analytic philosophy.

Having said this, while the 'Kantian tradition' and the contemporary analytic tradition have distinct methods of investigating imagination, the two traditions are not — and should not be — mutually exclusive. Mental states require cognitive faculties to produce them. Increasing our understanding of one aspect of imagination should help us increase our understanding of other aspects of imagination. As such, it is my hope that this Thesis indirectly contributes to progress in both fields.

1.2.3 Philosophy of scientific thought experiments

In philosophy of science, imagination has long been dismissed as an uninteresting 'psychological' concept. As Levy and Godfrey-Smith (2020, p.1) write:

> Science is both a creative endeavor and a highly regimented one. It involves surprising, sometimes unthinkably novel ideas, along with meticulous exploration and the careful exclusion of alternatives. At the heart of this productive tension stands a human capacity typically called "the imagination": our ability, indeed our inclination, to think up new ideas, situations, and scenarios and to explore their contents and consequences in the mind's eye.
>
> Despite its centrality, the imagination has rarely received systematic attention in philosophy of science. This neglect can be attributed in part to the influence of a well-known distinction between the *context of discovery* and the *context of justification* (Reichenbach, 1938),

> and a tendency in [logical] positivist and post-positivist philosophy of science to set aside psychological aspects of the scientific process. That situation has now changed, and a growing literature in the philosophy of science is devoted to the role and character of imagining within science.

One key topic that found itself in the limelight of philosophy of science in the past few decades is the topic of *thought experiments*. Galilei dropped balls off the tower of Pisa, Newton rotated a bucket of water in an empty universe, Maxwell conjured up a Demon, and Einstein rode on light beams and in space-bound elevators. Thought experiments are everywhere, both in the history of science and in contemporary science, often with far-reaching consequences. Initiated by the likes of Mach, Koyré, Popper and Kuhn, and with exponentially increasing effort since the 1990s, philosophers of science have set out to explain the ubiquitous presence and seemingly revolutionary capabilities of thought experiments in science.

Scientific thought experiments demand attention, first and foremost, because they are performed in the *imagination*. Scientists often seem to gain (scientific) knowledge or understanding about the world by performing thought experiments. But imagination is a highly controversial source of knowledge and understanding — certainly so in science. Science is a thoroughly *empirical* praxis: scientific knowledge and understanding must be gained by making *observations*, by gathering empirical data, by performing *experiments* in and on the world. Yet when we perform a *thought* experiment, we do not get new empirical data. When we perform thought experiments, it's just *us*. And yet we seem to learn about the world by performing thought experiments. And so, philosophers of science found themselves debating endlessly over the following question (Kuhn, 1977, p.241):

> How, relying exclusively upon familiar data, can a thought experiment lead to new knowledge or understanding of the world?

The past decades have seen a plethora of proposed philosophical *accounts of scientific thought experiments*. These accounts aim to describe

what scientific thought experiments are and what — and, more importantly, *how* — we can learn by performing them. There is still no broad consensus on any such account. In this Thesis (Chapter 4), I shall propose yet another philosophical account of scientific thought experiments, one which exploits recent developments in closely-related topics of investigation in philosophy of science (notably, the so-called *fiction view of models*), and which aims to strike a balance between the most plausible accounts available in the literature.

But, before I do so, I must first deal with a more fundamental problem: *what in tarnation is imagination*?

Chapter 2

Explicating Imagination

2.1 Introduction

After an attempt to analyse the concept of imagination in his influential work on the foundations of fiction and the representational arts, *Mimesis as Make-Believe*, K.L. Walton submitted defeat (1990, p. 18):

> What is it to imagine? We have examined a number of dimensions along which imaginings can vary; shouldn't we now spell out what they have in common? — Yes, if we can. But I can't.

Twenty years earlier, P.F. Strawson (1970, p. 31) submitted that the uses of the words image, to imagine and imagination are too variegated for even a family resemblance analysis. Despite increased effort to understand imagination during the last few decades, M.T. Stuart soberly concludes that little progress has been made (2021, p. 1329):

> The imagination is one of the most distinctive and philosophically interesting cognitive powers that humans possess. It is also one of the least well understood.

Few philosophers have attempted to explicate imagination. The next best thing to do seems to list typical but not essential features, characteristics, or "dimensions" as Walton calls them, of imagination. I shall

nonetheless make a sustained attempt to explicate the different types of imagination. This is the primary aim of the current Chapter. This analytic project of explication will involve explicating allied concepts, notably perceiving, conceiving, supposition, counterfactual thought, visualisation, picturing, hallucinating and remembering, and will involve forging explicit logical connections between imagination and those allied concepts.

The relevance of this analytic project directly concerns four lively domains of philosophical discourse: (i) philosophy of science, specifically the concept of *scientific imagination* and the *fiction view of models*, which is an exciting new account of scientific modeling that places the concept of imagination center stage[17]; (ii) the on-going discourse on thought experiments (experiments run in the imagination), in philosophy of science, metaphysics and philosophical methodology[18]; (iii) epistemology, specifically the issue whether imagination is a source of knowledge[19]; and (iv) aesthetics, e.g. Walton's (1990). My results can also be taken as part of a logical clean-up job of Walton (1990), thereby clarifying *its* conceptual framework. I specifically mention Walton's treatise because it grounds the fiction view of models (i) and, in Chapter 4, I shall employ the fiction view of models and the results of the current Chapter to analyse scientific thought experiments (i)–(iii).

Here follows a brief overview of what is coming. In Section 2.2, I provide the conceptual basis of my analytic project. In Section 2.3, I acknowledge a divide between contemporary philosophers thinking about imagination, I distinguish three types of imagination — proposition-imagination, entity-imagination and action-imagination — and I split entity-imagination into a sensory and a conceptual subtype. In Section 2.4, I work my way towards an explication of proposition-imagination, after which I go on to inquire whether my explication has features that other philosophers have ascribed to imagination. In Section 2.5, I explicate the allied concepts

[17] E.g. Frigg (2010b); Toon (2012); Levy (2015); Frigg and Nguyen (2016); Levy and Godfrey-Smith (2020); Salis (2021). [18] E.g. Nersessian (1993); Norton (2004b); Brown (2004); Gendler (2004); Meynell (2014); Stuart et al. (2018). [19] E.g. Kind and Kung (2016); Kind (2018); Kinberg and Levy (2022).

of supposition, counterfactual thought, conceiving, visualisation and picturing. In Section 2.6, I then distinguish two types of action-imagination, explicate them, and turn to the relation between imagination and memory. In Section 2.7, I recapitulate the explications proposed in this Chapter to provide an overview of what has been achieved (Figure 2.1) and I put forward my explication for imagination proper. Finally, in Section 2.8, I discuss a few phenomena that imagination exhibits in practice but which remained under-illuminated by the preceding conceptual analysis, and I make a few comments on the cognitive-scientific perspective on imagination, which will be relevant for the next Chapter.

2.2 Conceptual basis

2.2.1 Methodological preliminaries

In every project that analyses 'mental' concepts that adhere to a 1st-person perspective, like imagining, supposing, conceiving, picturing and visualising, the trouble is that many such concepts are not overtly and consistently connected to distinguishing observable behaviour that can be judged from a 3rd-person perspective.[20] To wit, what are the observable differences in behaviour of someone who imagines-that-p and someone who conceives-that-p? Between to-consider-that-p and to-suppose-that-p? Between to-entertain-p, to-attend-to-p, and to-think-of-p? If there are any observable differences at all, they lie predominantly in *linguistic* behaviour — that is, in speaking and writing.

But due to the context-dependent and individual subject-dependent usage of mental concepts, and in addition due to the unsurveyability of total usage, what any subject can come to know of linguistic behaviour underdetermines any analysis. Therefore *linguistic analysis* sooner or later morphs into what is currently called *conceptual engineering*: crafting con-

[20] This problem has haunted phenomenology from its inception; c.f. Philipse (2003). See however Section 2.8.2 in this Chapter for comments on the ways in which imagination *is* connected to observable behavior.

cepts for specific purposes, while covering *most* usage of the concept, rather than attempting to provide definitions of a concept that aim cover its usage universally.

The very idea of conceptual engineering is in perfect harmony with Carnap's (1950) celebrated conception of *explication*: to provide a criterion (necessary and sufficient conditions) for some concept (the explicandum) in terms of concepts that we understand better than the explicandum (yielding an explicans). For the explicans, Carnap submits, we must strike a balance between:

- ★ **similarity:** the explicans should cover a significant part of the usage of the linguistic expression of the explicandum;
- ★ **simplicity:** the explicans should include not too many concepts and these concepts should not cry out for analysis as loudly as the explicandum does;
- ★ **exactness:** the explicans should be an explicit logical combination of other concepts;
- ★ **usefulness:** the explication must serve a specified purpose.

When one explicates multiple concepts concerning one single topic, I add a fifth and sixth requirement:

- ★ **coherence:** logical relations between the explicated concepts should be straightforward to determine;
- ★ **consistency:** all explications taken together should form a consistent whole; no two explications should contradict each other.

I emphasise that there are many different, competing and often *incompatible* accounts available in the literature about every single concept that I explicate: entire monographs have been written about nearly every concept that I discuss and explicate in this Chapter. My proposed explications will never be in agreement with *all* these accounts. This would be impossible to achieve. But it is not my aim to propose explications that are in agreement with *all* accounts available in the literature. My aim,

rather, is to create a coherent and consistent *network* of concepts which are undeniably closely related but which have never been explicitly related to each other to the extent that I do in this Chapter. I aim to provide, in other words, a *conceptual geography of imagination and allied concepts*. In creating this conceptual geography, I aimed to strike a balance in my explications between (i) the above-mentioned six Carnapian requirements for explications and (ii) which ideas seem supported by *most* authors on the concept under investigation.

2.2.2 Results

For the sake of clarity, I next present the main results of my analytic project in a schematic overview: see Figure 2.1 (next page).

I first distinguish *imagination* from *perception*, *optical illusion* and *hallucination*. Within the concept of imagination, I then distinguish between imagining a *proposition* (proposition-imagination) and imagining an *action* (action-imagination). I then propose an explication of proposition-imagination, which result is the cornerstone of the conceptual geography of imagination presented in this Chapter. Using this explication of proposition-imagination, I then propose explications of *conceiving*, *supposition* and *counterfactual thought* as *types* of proposition-imagination. After this, I also explicate what it means to *visualise* and *picture* propositions. I then turn to explicating action-imagination. I distinguish between two types of action-imagination: imagining action from the *inside* and imagining action from the *outside*. I explicate each similarly. I then explicate what it means to visualise and picture actions. Finally, I relate the concept of imagination to *memory*, which relation turns out to be a rather surprising one.

A conceptual geography of imagination like the one presented in this Chapter has not been proposed in the literature before. By creating this network, I hope to increase the quality and clarity of future discussions about these concepts and their inter-relations. More directly, I make possible the unambiguous use of these concepts throughout this Thesis.

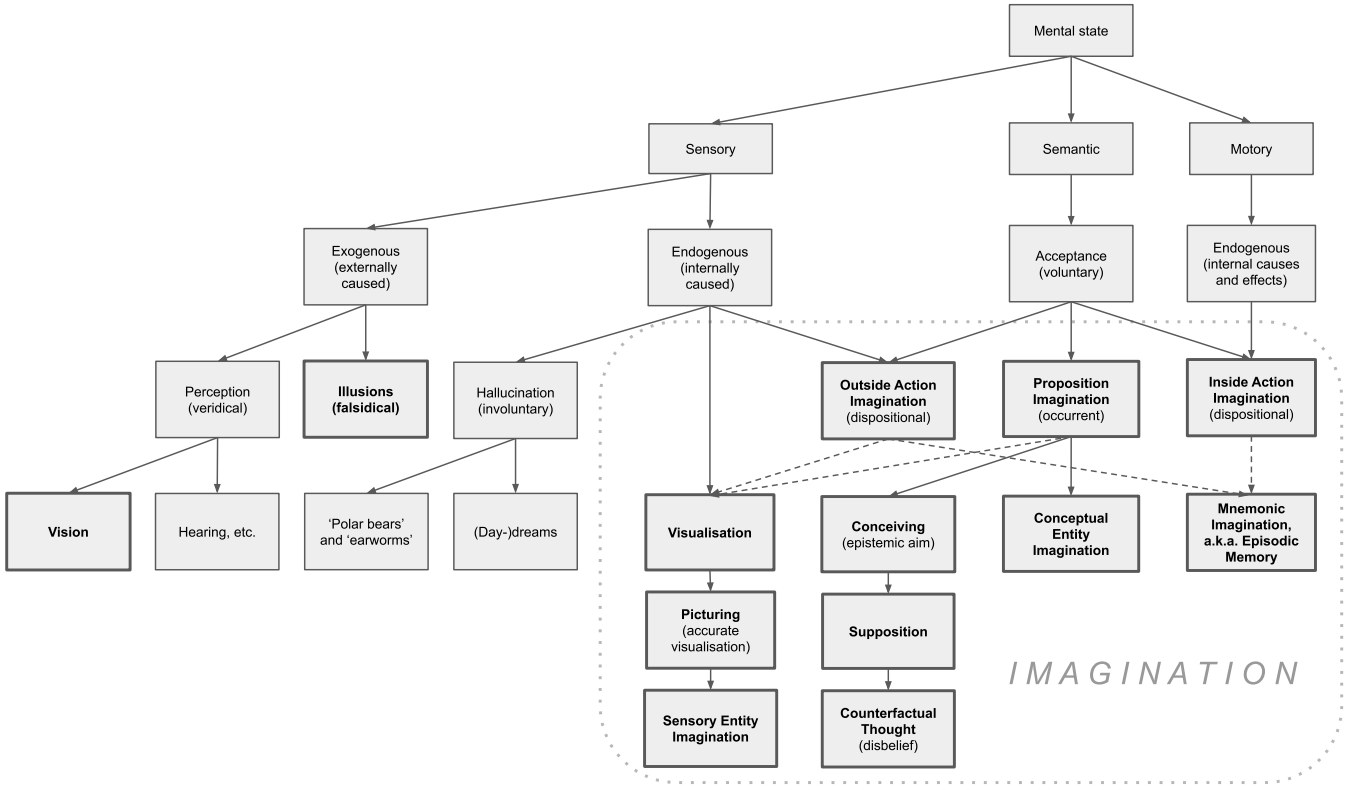

Figure 2.1: Diagram of logical relations between core concepts discussed in this Chapter: a *conceptual geography of imagination and allied concepts*. Arrows indicate necessary conditions; a set of dashed arrows indicate that only one member is necessary. I proposed explications for bold-faced concepts.

2.2.3 Conceptual basis

In every project of conceptual engineering and Carnapian explication, some concepts have to be taken for granted. They form the *conceptual basis* for that project. The members of my conceptual basis come from many branches of philosophy and are all pretty standard. I next mention them in order to be explicit about my conceptual basis.

Philosophy of Mind. I regard physical and mental (or psychological) as descriptive predicates, without making further metaphysical commitments; and I take heed of the just warning by Bennet and Hacker (2003, pp. 117–118) that the physical-mental distinction is neither a dichotomy nor always helpful.

I shall speak of *mental states* of subjects, their *content*, and their *intentional objects*.[21] An intentional object can be anything, just like the grammatical subject of a sentence can be anything. Relevant types of content are sensory, motor, semantic, mnemonic and affective. Sensory content is visual, auditory, tactile, olfactory or gustatory, corresponding to the five 'traditional' sense modalities. Motor (kinaesthetic, proprioceptive) content is not grounded in our sensory organs but in our "motor system (the motor and premotor cortices)" (Nanay, 2021, §4.1). One's mental state having motor content amounts to having the "feeling of doing something [...] from a first-person perpective" (*ibid.*): a feeling of muscles working, embodiment, agency, and explicit self-involvement. Semantic content of mental states is conceptual, as when one thinks of (some meaning of) words, or propositional, in which case I speak, following Russell, of *propositional attitudes*. I distinguish and frequently speak of *occurrent* and *dispositional* propositional attitudes; dispositional attitudes become occurrent when their manifestation conditions are met.

A mental state can have different types of content jointly, like a body can wear different garments jointly: Shelby remembered seeing and hearing an enormous whale (sensory: visual and auditory, mnemonic); Sinead

[21] The intentional object is not a relatum of some genuine relation between it and the content of a mental state, but rather a feature of the content worthy of attention (Ryle, 1971, p. 182).

angrily believes *that* the Pope is an evil lizard (semantic, sensory, affective); Sumaya fearsomely hears *that* a mouse is trotting around her bed (semantic, affective, sensory: auditory).

I explicate visual perception (a.k.a. *seeing*) and observation as follows:

> [Visual Perception] Subject S *visually perceives* concrete observable entity ε iff S has an occurrent mental state with visual content that represents ε, and has ε as its intentional object, and the subvenient brain state is externally caused by events that involve ε via an image of ε on the retina of the eyes of S and via signals in the optical nerve of S from eye to brain. (2.1)
>
> [Observation] Subject S *observes* concrete entity ε iff S deliberately and attentively visually perceives ε.

Similar explications for the other sensory modalities (hearing, etc.) I do not spell out; the explication of S *perceives* ε is the inclusive disjunction of all the explications of perception by all different sense modalities.

Call a perception of entity ε *veridical* iff it yields *true* perceptual beliefs about ε, and *falsidical* iff it does not.[22] Optical illusions are falsidical perceptions, for they induce false perceptual beliefs (Ebbinghaus' figure, Müller-Lyer's arrow, Ponzo's railway, Poggendorf's lines) or conflicting perceptual beliefs (duck-rabbit, young wife/old mother in law, Rubin's vase/face, Kanisza's triangle). See Figure 2.2 for examples. Explication:

> [Optical Illusion] Subject S has an *optical illusion* about (observable concrete) entity ε iff S sees ε and this induces either conflicting perceptual beliefs about ε (*ambiguous illusion*) or false perceptual beliefs about ε (*misleading illusion*). (2.2)

Next, following Langland-Hassan (2020, p. 78), I distinguish between *exogenous* sensory mental states, whose subvening brain states are exter-

[22] *Gratia* Quine (1962, p. 84) for the word falsidical.

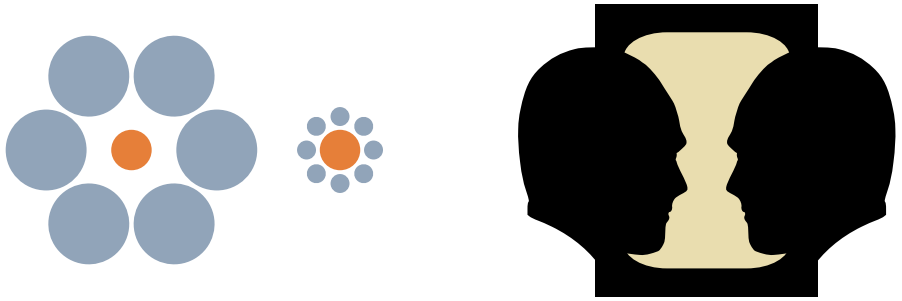

Figure 2.2: Ebbinghaus' figure (left) and Rubin's vase/face (right). (Both images are courtesy of www.wikipedia.com.)

nally caused by events via our sense organs, e.g. as in vision (2.1), and *endogenous* sensory mental states, which are mental states whose subvening brain state is *not* exogenous. Endogenous sensory mental states may be *voluntary*, or *involuntary*, which is a tricky distinction I shall elaborate upon when needed. Optical illusions and ordinary states of perception are exogenous sensory mental states. Mental states of imagination with sensory content I shall treat as *voluntary* endogenous sensory mental states. *Involuntary* endogenous sensory mental states I shall call *hallucinations*, see Section 2.4.1 for elaboration.

The sensory content of endogenous mental states — both voluntary and involuntary — is commonly referred to as mental imagery, which is a term that I occasionally employ but prefer to avoid. I limit myself here to a few short but necessary comments about this concept. Mental states with mental imagery are characterised by a "voluntary experience of creating a conscious sensory experience" (Pearson, 2019, p.624); c.f. (Nanay, 2021). Importantly, as was famously argued by Kosslyn (1980), mental imagery has besides semantic properties (content, reference, etc.) also *spatial* properties, or at least a functional analogue thereof. I emphasise that the "idea that pictorial representations [i.e. depictive mental imagery] are literally *pictures* in the head is not taken seriously by proponents of the pictorial view of imagery (see, e.g., Kosslyn and Pomerantz (1977)).

The claim is, rather, that mental images represent[23] in a way that can be functionally *like* the way pictures represent," e.g. in the sense that spatial distances between parts of the mental image are defined "in terms of the number of discrete computational steps required to combine stored information about them" (Pitt, 2022, §5). I return to the notion of mental imagery in Section 2.8.

Philosophy of Reality. I follow Lowe (2019, p. 49) by using the term *entity* to cover anything that belongs to some *ontic category*, which is a category of anything that does or can exist: concrete objects, abstract objects, facts, events, situations, structures, processes, contexts, worlds, and what have you. For the sake of brevity, I shall speak of existence of entities, although other words are appropriate when it comes to ontic categories other than material objects: facts obtain, events occur, situations are realised, processes unfold.

I point out that abstract entities include, besides mathematical objects like numbers, structures, sets and the like, also propositions and concepts (*salute* Frege). Propositions being entities falsely suggests that proposition-imagination is subtype of entity imagination. It is not: by 'imagining proposition p' I mean 'imagining *that* p', which denotes a mental state of imagination with semantic content p.

I shall avail myself of the concept of *truth-making*, with an all-inclusive take on truth-makers: any entity or plurality of entities can make a proposition true. Notably, mental states can be truth-makers, e.g. Stephanie's occurrent affective mental state of feeling pain makes true the proposition *that* Stephanie is in pain.

Philosophy of Language. For the sake of expediency, I adopt the Fregean ideas that predicates express concepts and sentences express propositions. Predicates have meaning-conditions, which tell us which categories of entities the predicate applies to in order to yield a meaningful linguistic whole, e.g. a *sentence*. Declarative sentences have truth-conditions, which tell us *which* entities must obtain in reality, and *how*, for the sentence

[23] See below for more on (scientific) *representation*.

and the expressed proposition to be true. I follow Berto (2022, Ch. 2) by analysing a proposition p in two components: the Truth-Conditions of p, abbreviated as: $\mathrm{TrC}(p)$, and the *topic*, or subject-matter, of p, which is *what p is about*, abbreviated as: $\varepsilon(p)$. This topic can be any entity or plurality of entities, even a proposition (Yablo, 2014; Berto, 2022). We have:

$$\text{Proposition } p \text{ is true in world } W \text{ iff } \mathrm{TrC}(p) \text{ obtains in } W. \tag{2.3}$$

The topic $\varepsilon(p)$ exists in W, unless p is or entails a negative existential about $\varepsilon(p)$. Although world-talk is entrenched in metaphysics, epistemology and modal logic, I want to emphasise that world can be replaced nearly everywhere in the current Chapter innocuously with: situation, scenario, environment, circumstance, context, etc.

Philosophy of Action. Actions are instances of behaviour, and behaviour is movement of the body with an intention or purpose. Behaviour is observable, and actions are observable events, as Davidson has taught us. This makes the concept of *mental action* an ostensible oxymoron, which is undesirable. Just as we can fulfil a request by an action, e.g. Would you pour me some of that Cabernet Sauvignon, please?, we can fulfil purely 'mental requests':

- ✔ Recite the alphabet in your mind.
- ✔ Picture yourself on a boat on a river.
- ✔ Imagine all the people, living life in peace.
- ✔ Think of a number between 0 and 100.

These requests make sense and can be fulfilled, or so I submit; they are requests for what I shall call *mental action*. By contrast, I shall call instances of behaviour, which are observable, *physical actions*. The oxymoronic character of mental action has hereby been exorcised.

Philosophy of Knowledge. The verb to think is a polysemous blunderbuss. Bennet and Hacker (2003, § 2) list *eleven* different meanings; I set

two of them apart: to *think-of* and to *think-that*. *To think-of* p seems to express the very same concept as several other English phrases, such as: entertaining, grasping, paying attention to, concentrating on, and having the thought that p. In contrast, *to think-that* p is markedly different from *to think-of* p, although if you think-that p, then you also think-of p. *Thinking-of* is closely connected to imagination, as I shall see; *thinking-that* is, by contrast, tightly connected to other epistemic concepts, as we shall see below.

I follow several philosophers by distinguishing sharply between the propositional attitudes of belief and of acceptance.[24] To *accept* that p is to take p *voluntarily* as a basis for mental or physical action, or both. It is making a *commitment to do something with* p whilst suspending epistemic judgment about p: to draw consequences from p, to relate p to what one knows or believes, to defend p when criticised, to use p when solving problems, to construct explanation involving p, to engage in deliberation on the basis of p, and so on, on the basis of what one already knows. In other words: to accept that p is voluntarily to think-that p is a basis for mental or physical action, or both, in many ways. Stalnaker (1984, p. 77) submits that while accepting that p, one ignores the possibility that p is false. By contrast, to *believe* that p is to think-that p is true, full stop. Belief is often involuntary, as in cases of perceptual belief. *Occurrent* acceptance, by contrast, is always the result of a conscious decision. Acceptance and belief neither exclude each other nor do they always walk hand in hand: you may consistently accept that p while believing that p or disbelieving that p.[25]

Finally, I take *knowing* that p for granted, and its implications of

[24] E.g. van Fraassen (1980); Stalnaker (1984); Cohen (1989); Engel (1998); Tuomela (2000). I do not distinguish different *types* or *modes* of acceptance. I do so mainly for practical reasons: this thesis is already long enough. I note that e.g. Arcangeli (2019) suggests analysing and comparing imagination and supposition to different *types* of imagination. Although I disagree with Arcangeli's preliminary conclusions from her initial exploration into this topic, I do agree with Arcangeli that this is an important area for future research. [25] van Fraassen (1980) connects *acceptance* of a scientific theory to *believing* only what it says about observables as the quintessential constructive-empiricist imperative.

truth and justified belief; nothing will hinge on what further is adduced to expulse Gettier. More on this in the next Chapter.

Philosophy of Science. Over the past decades, the concept of *representation* has occupied center stage in philosophy of science and has been strenuously debated. I neither want nor need to commit ourselves to any specific account of representation, because the use I shall make of it only requires a few basic features, which, as far as I know, all accounts share. The first feature is that representation is, or includes, a binary relation between typically an artifact (the representans: hypothesis, model, mathematical structure, theory, picture), which is an entity, and another entity, plurality of entities or *class* of entities (the representandum, the target). I write includes because there are several accounts of representation that do not take it to be a binary relation, with Van Fraassen as the champion by propounding a sexary relation, $\text{Repr}(S, V, A, \alpha, F, P)$: subject S is V-ing artefact A to represent target entity α as an F for purpose P.[26] (This sexary relation illustrates an important moral that is often obscured by treating representation as a binary relation: entities do not represent other entities 'by themselves', *people use* entities to represent other entities.) The second feature is that the representation-relation is irreflexive (an entity cannot represent itself) and anti-symmetric (if a represents b, then b does not represent a). The third feature is that representation is gradual in that there is a spectrum between *accurate* and *inaccurate* representation — 'more accurate' representation typically, but not necessarily, implies more correlated properties and relations between representans and representandum; c.f. Frigg and Nguyen (2022).

This concludes my conceptual basis. I move on to the typology of imagination.

[26] Four existential quantifications recover binary relation $\text{Repr}(A, \alpha)$; Muller (2009).

2.3 Typology of imagination

2.3.1 The Divide

To begin, I must acknowledge a deep divide among contemporary philosophers thinking about imagination. At one side of the divide we find *Imagers*, which are philosophers who want to keep *imagination* tied to its etymological root, the Latin *imago* (image):

> [Imagers] Mental states of imagination necessarily have sensory content. (2.4)

For Imagers, all mental states of imagination have endogenous sensory content, a.k.a. *mental imagery*. Notable Imagers are Wittgenstein (1980) and (Kind, 2001) — and nearly all philosophers from before 1900, recall the brief history of imagination in Chapter 1, Section 1.2. I emphasise that Imagers hold that *any* kind of endogenous sensory content will do, including non-visual sensory content (auditory, tactile, etc.): close your eyes and imagine the *sound* of a mosquito or the *feeling* of a hairy spider crawling over your face. I emphasise moreover that Imagers can (or should) acknowledge that a mental state of imagination can have *other* types of content besides sensory content; e.g. for Imagers, imagining a proposition yields a mental state with sensory *and* semantic content.

At the other side of the divide, we find *Wideheads*:

> [Wideheads] Mental states of imagination do not necessarily have sensory content. (2.5)

For example, Wideheads Bennet and Hacker (2003, p. 183) write:

> A powerful imagination is not the ability to conjure up vivid mental images, but rather the ability to think of ingenious, unusual, detailed, hitherto undreamt of possibilities.

Other notable Wideheads are White (1990); Walton (1990); Yablo (1993),

and, indeed, *most* contemporary authors on imagination.[27] Wideheads acknowledge that imagination *may*, and perhaps frequently *will* yield mental states with sensory content, they just hold that sensory content is not necessary (and not sufficient, for that matter) for exercising the cogitative capacity of imagination. I am a Widehead and shall propose an explication bereft of necessarily involving sensory content, but I shall keep track of how Imagers can adopt my explications.

2.3.2 Types of imagination

Imagining obviously is different from *knowing*, but there is an interesting parallel between the two. Regarding knowledge, Russell distinguished knowledge by description, a.k.a. propositional knowledge or knowledge-that-p, and knowledge by acquaintance, which is knowledge of entities gained mostly by means of direct sensory experience. Ryle added knowledge of doing something, a.k.a. practical knowledge or knowing-how-to-ϕ, as a third type. Then the question obtrudes whether there similarly are three types of imagination. Walton answers in the affirmative:[28]

> I postpone consideration of the differences between imagining a *proposition* [ImProp], imagining a *thing* [ImEnt], and imagining *doing something* [ImAct] — between, for instance, imagining that there is a bear, imagining a bear, and imagining seeing a bear.

In an oft-cited passage, Yablo (1993, p. 13) too distinguishes between ImProp and ImEnt (but not ImAct):

> Imagining can be either *propositional* [ImProp] — imagining that there is a tiger behind the curtain — or *objectual* [ImEnt] — imagining the tiger itself. [. . .] To be sure, in imagining the tiger, I imagine it as endowed with certain properties, such as sitting behind the curtain or preparing to leap; and I may also imagine *that* it has those

[27] E.g. Nichols (2009); Dokic and Arcangeli (2015); Salis and Frigg (2020); Langland-Hassan (2020); Berto (2022). [28] I have replaced Walton's '(i)' etc. with my acronyms.

> properties. So objectual imagining has in some cases a propositional accompaniment.[29] Still the two kinds of imagining are distinct, for only the second has alethic content — the kind that can be evaluated as true or false — and only the first has referential content — the kind that purports to depict an object.

The distinction between proposition-imagination (ImProp) and entity-imagination (ImEnt) is widely adopted and discussed in contemporary literature.[30] Salis and Frigg (2020) adopt this distinction as a basis of their typology of scientific imagination, which is one of the very few elaborate typologies of imagination on offer. Meynell (2021, p. 2) argued that the typology of Salis and Frigg is "both unmotivated and unconvincing". I agree with Meynell and aim to propose and motivate a richer and improved typology (see Figure 2.1 in Section 2.7 for results).

The concept of *action-imagination* (ImAct), mentioned by Walton, has been less often discussed than ImProp and ImEnt; it has been recognised by many but is rarely analysed or explicated.[31] This is problematic, because e.g. in many philosophical and scientific thought experiments, to imagine *doing* something is of the essence.[32] ImAct is "imagination in a richer sense, [as when] we immerse ourselves in a scenario, trying to 'live it' in our minds", as Kinberg and Levy (2022, p.3) put it, to simulate experiencing performing an activity. These phenomenological features suggest that there is a difference in the *contents* of the mental states of, on the one hand, ImAct, and on the other hand ImProp and ImEnt: ImProp involves predominantly *semantic* content (it is a propositional attitude), while ImAct involves predominantly *sensory* and *motor* content, and even *affective* content. On top of this, and perhaps unsurprisingly, the distinction between ImAct and ImProp seems grounded in different parts

[29] I return to Yablo's notion of *accompaniment* in Section 2.8.1. [30] See e.g. Nichols (2009); Kind and Kung (2016); Levy and Godfrey-Smith (2020); Salis and Frigg (2020); Liao and Gendler (2020); Berto (2022). [31] Notable exceptions are Dokic and Arcangeli (2015); Balcerak Jackson (2016), where Peacocke, Vendler, Gendler, Goldman, Curry and Ravenscroft are listed as reckoners. Goldman baptised it: *enactment-imagination*; Kinberg and Levy (2022): *immersive imagination*. [32] *Pace* e.g. Gooding (1992, 1993); Nersessian (1993).

of the brain (Nanay, 2021). Furthermore, there is a parallel distinction between propositional and action ('episodic') *memory* (Michaelian and Sutton, 2017), which is very relevant because memory is intimately tied to imagination (Section 2.6.3). ImAct must have a place in the analysis of imagination: in Section 2.6, I give it a place.

I reject however the categorical trichtomy between ImProp, ImEnt and ImAct, in that numerous mental states qualify as falling under several if not all of these categories. Notably ImEnt appears to reduce to ImProp and ImAct. I next consider ImEnt in some detail.

2.3.3 Reduction of entity-imagination

Salis and Frigg (2020, p. 27) describe ImEnt ("objectual imagination") as follows:

> Objectual imagination is a mental relation to a representation of a real or nonexistent entity. One can imagine London or the fictional city Macondo, Napoleon or Raskolnikov, a tiger or a unicorn. Yablo characterises objectual imagination as having referential content of the kind "that purports to depict an object" (1993, p. 27). Yet he emphasises that depicting an object does not require forming a mental image of it, which is why we can imagine objects that are hard (or even impossible) to visualise.

According to Yablo, and Salis and Frigg, imagining an entity does not require mental imagery. But what, then, *does* it require? What are we doing (if we can) when we are in a mental state without sensory content yet we are imagining an entity?

Since we obviously can, and often do imagine entities visually, and hence sensorily, I proceed by drawing first a provisional distinction within ImEnt, between: (*A*) ImEnt with mental imagery, and (*B*) ImEnt without mental imagery. I discuss each in turn.

(*A*) *ImEnt with Mental Imagery (ImEntSens).* I note the reasonably widespread agreement that "mental imagery is sensory imagination" (Ar-

cangeli, 2019, p. 19): to *imagine* an entity is the same as to imagine *perceiving* that entity.[33] As Smith (2006, footnote 18) writes: "We imagine a tiger *by* imagining seeing it." Since to perceive is an action, ImEntSens is an instance of action-imagination (ImAct): imagining *perceiving* something is imagining *doing* something. This conveniently reduces *sensory entity-imagination* (ImEntSens) to ImAct:

> [ImEntSens] Subject S *sensorily imagines entity* ε iff S imagines ϕ-ing, where ϕ-ing is the act of perceiving ε, that is, S imagines perceiving ε. (2.6)

When I have explicated ImAct (Section 2.6), I can substitute *its* explicans in (2.6); I have then explicated ImEntSens too.

(*B*) *ImEnt without Mental Imagery.* The only example of ImEnt *without* mental imagery that Salis and Frigg discuss is so-called *conceptual imagination*: to imagine an entity conceptually (ImEntConc). Salis and Frigg (2020, p. 27) continue after the above-quoted passage:

> However, if we cannot form a mental image of a chiliagon, how can we imagine it without imagining *that* it is so-and-so [ImProp]? Yablo does not consider this issue, but Gaut [2003] offers a natural solution: "Imagining some object x is a matter of entertaining the concept of x, where entertaining the concept of x is a matter of thinking of x without commitment to the existence (or nonexistence) of x" (2003,

[33] I acknowledge here a nitty-gritty debate about the content of mental imagery, between the so-called 'Dependency Thesis' (which holds that imagining an ε is imagining a perceptual experience of ε 'from the inside', i.e. with sense of agency or embodiment) and the 'Similar Content View' (which holds that imagining an ε is not necessarily imagining a perceptual experience of ε 'from the inside', it is just to have a mental state with the same content as a perception of ε); c.f. Peacocke (1985); Martin (2002); Currie and Ravenscoft (2002); Noordhof (2002); Nanay (2015). Although the details of this debate are beyond the scope of this Chapter, I am inclined to side with the 'Similar Content View': within action-imagination (ImAct), I shall distinguish sharply between imagining action from the inside and imagining action from the outside (Section 2.6), and I equate sensory entity imagination with imagining perceiving an entity from the *outside*, but not (necessarily) from the inside, as the Dependency Thesis purports.

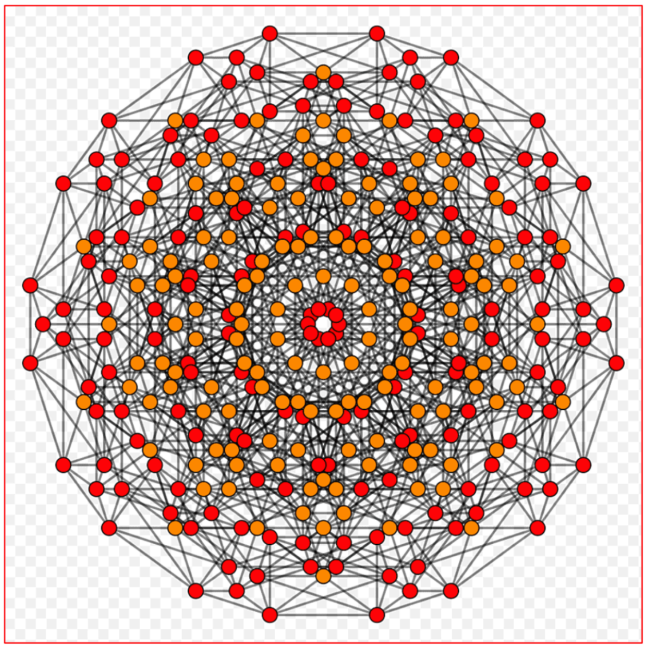

Figure 2.3: Visualisation of a chiliagon (courtesy of www.pngwing.com).

> p. 153). Imagining a chiliagon simply amounts to entertaining the concept of a chiliagon.

Before I continue, I want to flag that I frown at the claim of Salis and Frigg (and, famously, Descartes) that they cannot visualise a chiliagon. Granted, we cannot imagine an entire chiliagon in all its chiliagonic splendour, every one of its thousand sides jointly. But we cannot visualise or picture, and we even cannot see, all sides of a mountain jointly either. Should we now conclude that visualising mountains is impossible? That would be silly — we could even lay pictures of all sides of a mountain next to each other on the table: then we do see all sides of the mountain jointly. Equally silly is it to uphold that we cannot visualise a chiliagon. (See Section 2.5.4 for more on visualisation, and see Figure 2.3 for a visualisation of a chiliagon.)

Back to conceptual entity-imagination. I can formulate Gaut's explication as follows: Subject S *conceptually imagines* entity ε iff S thinks of a concept under which ε falls. When thinking of a concept is expressed by predicate C, then S thinks of $C(\varepsilon)$. Hence I arrive at:

> Subject S *conceptually imagines* entity ε iff S thinks of $C(\varepsilon)$, where C is a predicate expressing some concept. (2.7)

According to (2.7), conceptually imagining a chiliagon simply amounts to thinking of a thousand-sided polygon. But should we call such 'mere' thinking-of a predicated entity *imagination*? I think not. It is not true that, if one thinks of *anything*, then they imagine it. Imagination requires something *more* than thinking of a concept. I side with Walton (1990, p. 13), who explicitly argues that the concept of thinking-of (entertaining) is insufficient to capture imagination:[34]

> Occurrent imagining, as we ordinarily understand it and as we need to understand it in order to explain representation, involves more than just entertaining or considering or having in mind the propositions [or entity or action] imagined. Imagining, [like] believing or desiring, is *doing* something *with* a proposition one has in mind.

So, imagining an entity is more than just thinking of a concept under which the entity falls. Imagining involves *doing* something with the entity that one has in mind (see also the quote of Bennett and Hacker on p. 30). If this 'doing something' does *not* amount to imagining *perceiving* the entity (ImEntSens), which I reduced to ImAct (2.6), then surely it amounts to imagining *that* the entity is such-and-so (whatever *that* amounts to; see Section 2.4.1 for my proposal), as Salis and Frigg suggest themselves.

[34] Admittedly, Walton is after a type of imagination that is involved in our engagement with fiction, which is clearly an 'active', 'participative' form of imagination. However, the idea that imagination involves *doing something* with the entity (or action or proposition) that they imagine has been argued for more often; see notably Kind (2001), who explicitly argues that imagining is an *activity*.

Hence, an improvement on the Gautian explication (2.7) in my eyes is:

[ImEntConc] Subject S *conceptually imagines* entity ε iff S imagines *that* $C(\varepsilon)$, where predicate C expresses some concept. (2.8)

Hereby I have reduced conceptual ImEnt to ImProp. When I have explicated ImProp, I can subtitute *its* explicans in (2.8), and take for p the proposition *that* $C(\varepsilon)$.

I emphasise that neither explicans (2.8), nor (2.7) for that matter, *guarantee* that ImEntConc yields a mental state without sensory content, in spite of the idea that ImEntConc was supposed to be non-sensory. Imagine a tomato conceptually. Can you do it, without any sensory content? Seems hard. Whether or not imagining an entity falling under a concept yields a mental state without sensory content depends heavily on that very concept. Explication (2.8) of ImEntConc permits this; I have not and will not add a conjunct to the explicans expressing that the mental of ImEntConc has no sensory content. Imagine an inaccessible infinite cardinal conceptually. Perhaps you can do *this* without sensory content. If you can do it without mental imagery but imagine *that* the inaccessible infinite cardinal is such-and-so, then ImEntConc remains a subtype of ImProp, just as in (2.8).

To recapitulate, I have argued that: (i) to imagine an entity sensorily (ImEntSens) reduces to ImAct (2.6), because to imagine an entity sensorily is the same as to imagine *perceiving* that entity; and (ii) to imagine an entity conceptually (ImEntConc) reduces to ImProp (2.8), because to imagine an entity conceptually is the same as to imagine *that* the entity is such-and-so. I have no reason to believe there are more subtypes of ImEnt, so I take the subdivision to be exhaustive:

[ImEnt] Subject S imagines entity ε iff S imagines ε sensorily (2.6) or S imagines ε conceptually (2.8). (2.9)

So much for entity imagination. Next: proposition-imagination.

2.4 Proposition-imagination

2.4.1 Explicating proposition-imagination

It seems that we can imagine *any* proposition that we want. You can imagine that there is a tiger behind the curtain. You can imagine that quantum mechanics is false. You can imagine that the Earth is a perfect sphere, or that it is flat as a pancake. You can imagine that the Bohr model of the Hydrogen atom is correct, and imagine that the aether exists after all. You can imagine that you are Batman. You can imagine that *tertium non datur* fails, you can even imagine that some contradiction is true. You can imagine that you are living in a world of philosophical zombies. When it comes to imagination, the sky is the limit. As Einstein famously quipped: logic will get you from a to b, imagination will take you *everywhere*.

Let me make the idea that we can imagine *any* proposition more precise. Hume famously argued with his *recombination principle* (Chapter 1, Section 1.2) that *we can imagine every possible recombination* of ideas that we have previously had. Most contemporary authors go beyond Hume's recombination principle and argue that we can even imagine the *impossible*; see e.g. Walton (1990, pp. 32–34, 64–67); White (1990, pp. 179–183); Berto (2017); Berto (2022, §5.1) and references therein. In short, I shall proceed with the following assumption: if we understand the meaning of a proposition — whatever possibility or impossibility it expresses — then we can imagine it.

But when Sofia is imagining proposition p, what is her mental state?

Well, when Sofia is imagining that p, she has a mental state with semantic content p, so at the very least she is *thinking-of* p. But there is more to imagining p than merely thinking of p, as I argued in the previous Section. What, then, does imagining p amount to — what do we *do* with p when we imagine it? White (1990, p. 184), in his impressive but curiously overlooked monograph on imagination, connected imagination solely to

possibility:

[White] Subject S imagines that p iff S thinks-of p as possible. (2.10)

To connect imagination to possibility is universally acknowledged as a crucial step in the right direction. Bennett and Hacker (2003, p. 182) follow suit in calling imagination "the cogitative capacity to think of possibilities". Langland-Hassen (2020, p.95) argued that "imagining involves contemplating possibilities in a rich and epistemically safe way." The very first paragraph of the SEP-entry on imagination states that "[o]ne can use imagination to represent possibilities other than the actual" (Liao and Gendler, 2020, p.1). Even continental philosophers concur, see e.g. Aldea (2019). Thus any explication of ImProp should contain the modal concept of possibility.

In Section 2.2, I noted that thinking-of is a rock-bottom mental concept that is synonymous to many other English phrases, such as entertaining, grasping, paying-attention-to, having-a-thought. None of these phrases capture adequately that imagining a proposition is "*doing* something *with* a proposition one has in mind" (Walton, 1990). Let me try to find an alternative for thinking-of.

What about replacing thinking-of with the more familiar doxastic propositional attitude of belief? This would give:

S imagines that p iff S believes that p is possible. (2.11)

Many contemporary authors connect ImProp to belief: e.g. Kind and Kung (2016, p.3) describe it as "belief-like but not quite belief", Dokic and Arcangeli (2015, p.12) as "a re-creat[ion] of a conscious occurrent belief", and Langland-Hassan (2012) even argued that propositional imagination *is* a form of believing. Nevertheless I reject explication (2.11) for two reasons.

First, belief is typically involuntary, often the result of unconscious cognitive processing (e.g. perceptual beliefs) or of conscious deliberation

(e.g. complicated scientific hypotheses). Proposition-imagination is nearly always voluntary; one is nearly always free to decide, or to refuse, to imagine something. Nearly! In *Winter Notes and Summer Impressions* (1863), Fjodor Dostoevsky writes:

> Try to pose for yourself this task: *not* to think of a polar bear, and you will see that the cursed thing will come to mind every minute.

Psychologists Wegner and Schneider (2003) established such 'ironic processes': requesting test-persons *not* to imagine a polar bear made it far more likely they did imagine one. I shall deal with Dostoevsky's polar bear in Section 2.4.2 (feature II). I only note here that imagination is *far more often* voluntary than belief.

Secondly, belief is 'too strong'. Belief is aimed at truth, to believe that p is to think-that p is *true*. This is not the case for imagination. Sissy can imagine that an unstoppable force meets an immovable object. This is more like *thinking-of the possibility* that an unstoppable force meets an immovable object, while she does not *believe* that this is possible, because Sissy is convinced this is *impossible*. Imagining that p is neither believing that p nor believing that p is possible. We need a propositional attitude that is far more voluntary and more 'active' than 'passive' belief.

My choice is *acceptance*, as analysed by Cohen (1989) and others (fn. 24, Section 2.2; ; c.f. Arcangeli (2019).). Acceptance *is* the voluntary, 'active', distinct sibling of belief. To accept that p is to adopt p voluntarily as a basis for mental or physical action, or both. This coheres with the idea that imagining a proposition is *doing* something *with* that proposition, as Walton would have it, and it is moreover in perfect harmony with White's conclusion (1990, p.183) that:

> ... imaginability [is] both a sufficient and necessary test for the *acceptability* of something as possible.

Furthermore, proposition-imagination is occurrent — it is occurrently *doing something* with a proposition — so the *acceptance* needs to be *occur-*

rent rather than dispositional. This gives:

S imagines that p iff S occurrently accepts that p is possible. (2.12)

There are many types of possibility: e.g. logical, metaphysical, nomic, epistemic, technological, sociological, political, psychological, personal, quotidian, and what have you — which type do I have in mind? I do not believe that in the explication of ImProp I can and should commit to a specific type of modality. Instead, I hold that the to-be imagined proposition, and the context in which it is imagined, narrows down if not determines an *appropriate* modality type for the imagining subject. To illustrate, the appropriate modality type for imagining that there is an elephant in the room is practical possibility, rather than nomological or metaphysical possibility; for imagining that philosophical zombies exist, the type of possibility is metaphysical or epistemic, rather than technological or nomological.

Let τ be a type of modality, ranging at least over the types mentioned above. Then my proposed explication of proposition-imagination is:

[ImProp] S imagines that p iff S occurrently accepts that p is τ-possible, for some appropriate modality type τ. (2.13)

I next make two systematic comments about this explication.

First, with (2.13) I claim that the relevant propositional attitude for proposition-imagination is *acceptance*. This is quite an unorthodox suggestion. Although imagination-like mental states such as supposition and counterfactual thought have often been explicitly connected to acceptance (see Sections 2.5.1–2.5.3), it is not often explicitly argued that *imagination* itself entails the propositional attitude of acceptance. Throughout this Chapter, I shall argue in favor of explication (2.13) mostly by demonstration, i.e. by showing that this explication enables us to logically connect imagination to its related concepts in a remarkably straightforward and elegant way. I take my conceptual geography of imagination (Figure 2.1)

to *be* an argument in favor of my proposed explication (2.13). Still, it will be useful to sketch the motivation for this explication by looking at the biconditional in (2.13) from both sides.

(i) If S imagines that p, then S occurrently accepts that p is possible. I have already argued for this conditional claim throughout this Section. I mentioned, notably, that White (1990) noted that imagination is "both a sufficient and necessary test for the acceptability of something as possible." To see this, consider colloquial uses of imagination. Sam desperately asserts: I cannot imagine that someone will return your lost wallet. By this, Sam means: I cannot accept that it is (practically) possible that someone will return your lost wallet. (Note that, in order to understand one's statement about their *ability* or *inability* to imagine something, sensitivity to the (implicit) *appropriate* type of modality is crucial. If Sam says that he cannot imagine that someone will return your stolen wallet, then he cannot accept that it is *practically* possible that someone will return your stolen wallet. Of course, Sam *can* still accept that this is e.g. *nomologically* possible. But say that one can imagine something *per se*, i.e. in *some* type of modality, would be an uninteresting tautology: for every proposition p, there is always a modality such that p is possibly true.)[35] Shahid says: Imagine that you accelerate from rest to a speed larger than the speed of light. Hereby, Shahid asks you to accept that it is (metaphysically) possible that you so accelerate. The request to imagine is a request to *accept* some possibility — to voluntarily take some possibility in mind and treat it *as if* it were real.

(ii) If S occurrently accepts that p is possible, then S imagines that p. This conditional claim is perhaps more questionable than its converse. Notwithstanding, I insist on this conditional claim too. I emphasise that occurrent *acceptance*, by itself, does *not* imply imagination — only the occurrent acceptance of a *possibility* does. To see this, consider the following examples. Shirley accepts that there is a cat on the mat. No imagination implied here. But now Shirley accepts that it is practically *possible* that

[35] I thank Stefan Wintein for drawing my attention to this point.

there is a cat on the mat, even though there is no cat on the mat. (Perhaps Shirley accepts this possibility because her cat sadly passed away last year, and she is entertaining the thought of getting a new one.) What does Shirley do in this case? She *imagines* that there is a cat on the mat. Next, a more abstract example. Sabrina occurrently accepts that it is possible that the axioms of Euclidean geometry apply to some certain type of space. What does Sabrina imagine in this case? She imagines *that* the axioms of Euclidion geometry apply to that type of space — presumably, Sabrina even *supposes* it (Section 2.5.1).

I am presently unaware of any knock-down counter-examples to conditional (ii). But perhaps one still has the strong intuition that it is *not always* the case that, if one occurrently accepts that p is possible, then one imagines that p. That is, perhaps one has the intuition that one can accept that p is possible *without* imagining p. I respond that *conceptual engineering* comes to the fore here. By *insisting* on conditional (ii), all the pieces of the conceptual puzzle of logically connecting imagination to its allied concepts fall into place, as I shall demonstrate in the next Sections. This is the very aim of my current conceptual engineering project. Unless and until there are knock-down counter-examples against conditional (ii), I submit that *accepting* conditional (ii) is highly fruitful and worthwhile.

Second, a few notes on the concept of *(im)possibility* in explication (2.13). I have already argued that imagination should be connected to the concept of possibility — about this, there has been near-unanimous consensus at least since Hume. But I also noted that, according to broad contemporary consensus, even contradictions and 'impossibilities' are imaginable.[36] This raises the question: how does my proposed explication of proposition-imagination (2.13) deal with imagining impossibilities?

To begin, I note that imagining impossibilities is a highly subtle affair, and it is not at all clear how we *should* account for it. What do we *do* when we imagine impossibilities? Consider imagining Escher's impossible cube;

[36] Again, see e.g. Walton (1990, pp. 32–34, 64–67); White (1990, pp. 179–183); Berto (2017); Berto (2022, §5.1) and references therein.

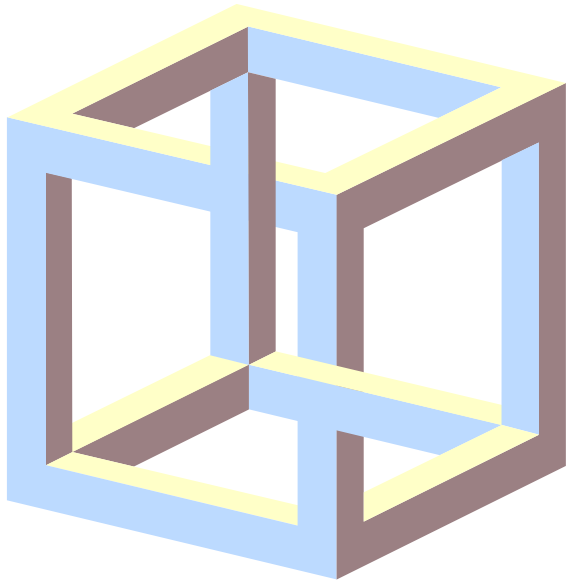

Figure 2.4: Escher's impossible cube. (Image courtesy of wikipedia.com.)

see Figure 2.4. When I try to imagine this cube, I find myself quite unable imagine the impossible cube as a whole. Instead, I find myself constantly *switching* between *conflicting* views of this cube as a *possible* cube. At one instant I imagine the second-left vertical leg of *behind* the second-right vertical leg, and the next instant I imagine it *in front of* the second-right vertical leg. In other words, I find myself imagining this impossible cube *as a possible cube* in various *conflicting* ways. This is, roughly speaking, White's (1990, pp.179–183)'s position too: we imagine impossibilities by thinking of them *as possibilities*.

But an answer along these lines is not the full story. To imagine conflicting possibilities is not quite to imagine an *impossibility* in the fullest sense. So, what would imagining Escher's impossible cube in its *entirety* amount to, i.e. how would we imagine that the second-left vertical leg of the cube is both *behind* and *in front of* the second-right vertical leg? My explication (2.13) provides an answer, by pointing our attention to the *appropriate* modality type wherein we accept the possibility of two

incompatible propositions being true at the same time. When we imagine impossibilities — i.e. an impossibility from the perspective of classical, bivalent logic —, the appropriate modality type will be provided by some paraconsistent logic, which makes some contradictions no longer false but turns them into possibilities. This parenthetically is another reason why acceptance appears to me as the appropriate propositional attitude: acceptance is not belief, so one might *accept* that possibly $q \wedge \neg q$ for some specific proposition q, for the sake of whatever, even if one does not *believe* that it is possible that $q \wedge \neg q$ is true.

After all this possibility-talk, it is perhaps desirable to have a more substantive account of possibility present in my explication. Substituting the standard account of possibility in terms of possible worlds[37] in explication (2.13), my explication of ImProp becomes:

> [ImPropWorld] Subject S *imagines that* p iff S occurrently accepts that there is a possible world where p is true, for some appropriate modality type τ. (2.14)

I shall use this explication in the remainder of this Chapter.

Let's next find out whether my proposal (2.14) fits in with recent ponderings on imagination.

2.4.2 Features of imagination

Below I list eight features which have been mentioned as *typical features of imagination* by various authors; see e.g. Nichols and Stich (2000); Nichols (2009); Gendler (2010); Liao and Gendler (2011); Dokic and Arcangeli (2015); Kind and Kung (2016); Frigg and Salis (2020); Liao and Gendler (2020); Badura and Kind (2021); Özgün and Schoonen (2022). I evaluate whether all eight features *should* be regarded typical features of imagination and I indicate whether, and if so, how, each feature is accounted for

[37] To account for imagining impossibilities, we should allow *impossible worlds* in our conception of 'possible worlds' too; c.f. Berto (2017); Berto and Jago (2019).

by my proposed explication (2.14).

I. *Episodic, Temporary.* Sydney imagines There is a cat on the mat and somewhat later she imagined The cat is chasing a mouse in the living room. She begins imagining, imagines for a while, and then stops imagining. Depending on the imagined proposition, imagining *that* p may consist of a single mental state and it may consist of a sequence of different mental states. In both cases Sydney begins imagining and somewhat later stops imagining it, which is to say that Sydney imagines during some episode. The episodic and temporary character is undeniably true for imagination, and it is in full agreement with *occurrent* acceptance (2.14).

II. *Voluntary, Deliberate.* Sally can choose *whether* to imagine There is an elephant in the room (α), and she can choose to some extent *how* to imagine α. Balcerak Jackson (2016, p. 45) defends for imagination "the Voluntary Control Thesis", which my explication (2.14) thus solemny obeys: acceptance is voluntary. The significant freedom of the proposition-imaginer to choose voluntarily *how* to imagine that α is accounted for by (2.14), because accepting *any* α-world from an infinitude of α-worlds will do (but only α-worlds will do).

I note that, although it is uncontroversial to claim that we can voluntarily choose the *topic* of our imagination, it remains a point of controversy to what extent we can determine the *content* of our imagination. Sally may choose to imagine the elephant *as* happy, but *how* exactly she will imagine the elephant as happy will be to a large extent involuntarily determined by subject-dependent factors such as background beliefs, memories, primes, expectations, etc.; see e.g. Langland-Hassan (2016), c.f. Section 2.8 of this Chapter. It seems safest to say that we can typically choose the *topic* of our imaginings, but that a choice of topic *under-determines* the *content* of our imaginings. This too is accounted for by my proposed explication (2.14): accepting *that* there is an α-world where p is true under-determines the α-world where p is true that one *actually* will have in mind. I will return to this point at Feature VIII. *Under-determination.*

This leaves me with the case of Dostoevsky's polar bear (Section 2.4.1).

We often imagine a polar bear involuntarily at the explicit request Do *not* imagine a polar bear. By sincerely attempting to obey, we disobey. Some mental imagery is *not* voluntary. Dostoevsky's polar bear shows the importance of conceptually distinguishing between: (i) *mental imagery*, i.e. the content of endogenous sensory mental states, which may be either voluntary or involuntary; and (ii) *imagination*, which is voluntary. Voluntary mental imagery is related to imagination, but is neither sufficient nor (according to Wideheads) necessary for it (c.f. Section 2.5.4 and 2.6). Mental states of imagination with voluntary mental imagery I call *visualisations* (Section 2.5.4) and, more generally, states of *action-imagination* (Section 2.6); by contrast, mental states with involuntary mental imagery I call *hallucinations*.[38] I next comment on the 'voluntariness' that distinguishes visualisations from hallucinations.

Both hallucinations and visualisations are falsidical endogenous sensory mental states. The distinguishing difference is voluntariness, which is a notion that must be handled with care: Susie may be aware that she is hallucinating, and she may even voluntarily choose to trigger hallucinations, e.g. by taking LSD; but Susie cannot choose *to stop* hallucinating after she has relished in LSD. This marks the difference with visual imagination: Susie can choose, at any moment, to begin visualising something, and to stop visualising it, but she cannot choose to hallucinate something; and after having taken LSD, she cannot choose to stop hallucinating either. Furthermore, Susie has little say in *what* she hallucinates, the content of her hallucinatory mental states is largely beyond her control. By contrast, visualising that p yields sensory mental states severely constrained by p.

Dostoevsky's polar bear involves involuntary mental imagery, hence it is a hallucination — a harmless 'gentle' hallucination, if you will. Similar examples of such gentle hallucinations are earworms, i.e. songs that involuntarily run continually through one's mind. At the risk of overstepping my boundaries, I think that day-dreams, dreams and nightmares also deserve to be called hallucinations. This I deem at least preferable over

[38] I here follow notably Nanay (2016a).

regarding dreams as instances of ImProp, like Salis and Frigg (2020) do.[39]

III. *Possibility.* Imagination involves possibility, no doubt about it. Recall e.g. that Hume set forth that "nothing we imagine is absolutely impossible" (1963, p. 32, 5.1), and that White (1990) argued that the only thing that all proposition-imaginings share is that they involve "thinking-of p as possible" (2.10). A fine-grained conception of possibility is prominently present in my explication of ImProp (2.14) and discussed throughout this Chapter.

IV. *Independence.* Perception and imagination are logically independent propositional attitudes. Sandra sees that there is a robbin in the garden; she is not imagining this. Sandra sees no elephant in the garage; she is however imagining that there is one. *Mutatis mutandis* for belief and imagination. Sandra does not believe that there is an elephant in the garage, but she can imagine that there is one. Occurrently accepting that p is possible (2.14) does neither imply believing nor perceiving that p is possible, nor *vice versa*. Hence explication (2.14) is logically independent of perception and belief.

Admittedly, subtle issues remain; I mention one. When Sandra occurrently perceives and believes that p, the *actuality* of p may be much more occurrent to her than its *possibility*. It is not obvious, then, that Sandra can *imagine* that p when she already perceives and believes that p. Suppose that Sandra is requested to look at a real elephant in the garage. Then, Sandra is asked to *imagine* this elephant in the garage. Strange question, because when you see something, you don't *have to* imagine it anymore — what would be the point? But *can* you not imagine that what you already perceive and believe? Answering in the negative would mean that perception and belief are not logically independent of imagination. An answer in the negative is however too quick. Sandra *can* imagine that what she occurrently perceives and believes already — every proposition

[39] Salis and Frigg (2020) seem to confuse the semantic content of a mental state and a description of the content of a mental state: these do not and cannot coincide when the mental state has no semantic content, which surely can be the case for dreams.

is always imaginable — if she puts in the effort to make the possibility of p 'more occurrent' than its actual truth, e.g. by first considering the possibility that p is false, then that it is possibly true. Let me illustrate this with an example.

Suppose that Sandra is looking at a finger on her hands that is without a ring. She then perceives and believes *that* this finger does not have a ring on it. Then, Sandra is asked to *imagine* that this finger does not have a ring on it. Sandra *can* do this, but she needs to 'trick' herself into it, as follows. First, Sandra looks at a ringless finger on her hand. Next she imagines that this finger *does* have a ring on it. Easy. Next Sandra imagines that she takes the imagined ring off her finger. Also easy. Finally, she imagines that this finger does *no longer* have a ring on it. Still easy this time. But there wasn't a ring on her finger in the first place and she saw and believed that! *Ta-da!* Sandra has imagined something that she sees, a ringless finger, which means that perception and imagination are logically compatible. This is in agreement with my explication of ImProp (2.14) and underwrites the logical independence of ImProp and belief.

V. *Quarantined.* Due to the fact that imagination is logically independent from belief and perception, imagination is largely quarantined from physical action and other typical consequences of beliefs and perceptions. Imagining that your house is on fire does not make you jump up in a hurried frenzy, grab your precious belongings, and run into the street. Whereas if you (perceptually) *believe* it, you will do precisely this.

Yet van Fraassen (1980) has taught us that acceptance *is* tied to action, which makes imagination according to my explication (2.13) not quarantined. Now what?

Gendler (2010, Chapters 7 and 12–14) has eloquently expounded that imagining that p *can* have "effects that we would expect only perceiving or believing that p to have" (2010, p. 238), that there can be "contagion" from imagination to mental and physical action. Imagining that your dog dies may actually make you experience sadness and cause you to go pet

your dog; imagining that your high-school crush is watching you perform a skateboarding trick (while she is absent) may actually make you try harder to land the trick. Hence quarantining from action is not a defining feature of imagination. Indeed, Nanay (2016b, p.127) "emphasizes the role imagination plays in our decisions: when we decide between two possible actions, we imagine ourselves in the situation that we imagine to be the outcome of these two actions and then compare these two imaginings." Williamson (2016) has even argued that imagination has the *primary aim* of providing practical knowledge, i.e. to guide our actions, like when you look at a cliff and imagine climbing it, and then climb it as imagined. See also Van Leeuwen (2016) for a nuanced account of the various ways in which imagination motivates action.

Explication (2.14) can account for this all. If for τ, the *practical* modality type is chosen, then the imagination likely is not excluded from action: the things we imagine in an imagined practical context relate directly to *actual* practical contexts. But pondering metaphysical possibilities that are not nomic possibilities will be — and better be — quarantined from physical action. Balcerak Jackson (2016, p. 44) holds that imagination is "epistemically innocent": imaginings cannot support beliefs. My response: depends on the appropriate type of modality. Epistemic innocence does not imply epistemic *impotence*; see Chapters 3 and 4 of this Thesis.

VI. *Belief-like.* Proposition-imagination has been described vaguely as "belief-like but not quite belief" (Kind and Kung, 2016, p.3), or even as the "re-creation of a conscious occurrent belief" (Dokic and Arcangeli, 2015, p.12). Further, it has been argued that ImProp is like belief in that both "exhibit[s] inferential orderliness" (Nichols, 2009, p. 365): the inferences that we make with imagined propositions 'mirror' or 'parallel' the inferences that we would make if we were to *believe* those propositions. Explication (2.14) accounts for this, because *accepting* a proposition involves, amongst other things, a commitment to making inferences with it that are appropriate in the relevant context (c.f. Section 2.2). Berto (2021, p. 2033) reports that the imagined proposition is integrated with

what the imaginer believes and knows. Accepting some p-world (2.14) entails precisely this.

Needless to say, imagination is also unlike belief in some respects. One notable difference is that belief aims at truth, one ought to believe the truth (Chignell, 2018), whereas imagination is not governed by such an imperative — and neither is acceptance.

VII. *Perception-like.* I have argued above, in feature IV. *Independence*, that perception and imagination are logically independent. *Typically*, however, imagination and perception are mutually exclusive in practice, as Wittgenstein (1980, p.13) expressed: "While I am looking at an object, I cannot imagine it." But in *some* cases they can walk hand in hand, as the example of Sandra imagining that there is no ring on her ringless finger showed us; this confirms their logical independence.

The likeness of imagination to perception is that a subspecies of perception, *visual* perception, has mental states with visual content in common with a subspecies of imagination, namely visualising and picturing, no more and no less. I analyse visualisation and picturing in Section 2.5.4, the results of which will underwrite the perception-likeness just mentioned. (For Wideheads, this should do. Imagers, however, will reject my explication of ImProp (2.14) simply because it does not require *sensory content*. For Imagers, then, to imagine a proposition is to *visualise* it.)

VIII. *Under-determination.* John Lennon sings a request to Yoko Ono: Picture yourself on a boat on a river, with tangerine trees and marmelade skies. Suppose that Yoko is now imagining *that* she is in a boat on a river, with tangerine trees and marmalade skies (α). Voluntarily and deliberately, Yoko occurrently accepts *that* α is possible, including *how* α is possible. To wit, she can choose the color, shape, size and type of the boat, she can choose any shade of orange for the tangerines, she can choose the size of the tree, she can choose a couple of trees or a forest, she can choose a brilliantly shining sun in the sky or marmalade clouds blocking the sunlight, she can choose to row the boat or let it gently flow down the stream, she can choose a calm river or a river wild, etc. With a slight variation on Berto

(2021, p. 2034): proposition p *under-determines* the content of the mental state of imagining that p. My explication (2.14) accounts for this under-determination because accepting *any* p-world suffices to imagine that p.

Relatedly, as I already mentioned in my discussion of feature II. *Voluntary, Deliberate*, voluntarily choosing a *topic* to imagine under-determines the *content* of our mental state of imagination. As Langland-Hassan (2016) puts it: the content of our imagination is not only determined by "top-down" rules. This problem is interestingly recursive. No matter *how specific* one chooses the topic of one's imagination, the content is always under-determined by this choice — the content of our imagination is under-determined all the way down. I discuss how this works *in practice* in Section 2.8 of this Chapter, and more extensively in Chapter 3. I here only repeat that my explication (2.14) can account for this all: occurrently accepting *that* there is a possible world where p is true under-determines *which* possible world where p is true (of which there are usually infinitely many) one actually has in mind.

To recapitulate, my explication of proposition-imagination (2.14) vindicates all features that have been noted as *typical features* for imagination by various authors (references at the beginning of this Section). I take this to be a big win for my proposed explication.

To demonstrate the usefulness of explication 2.14, I next use (2.14) to explicitly relate imagination to the allied concepts of (propositional) supposition, counterfactual thought, conceiving, visualisation and picturing. I provide explications of each concept in the process.

2.5 Allied concepts

2.5.1 Supposition

Supposition is the relevant propositional attitude in many philosophical and scientific thought experiments, in reasoning with possibilities other than the actual, and in reasoning through *reductio ad absurdum* arguments: suppose that a trolley charges down a railway track and you are

faced with a choice; suppose that you are in charge of developing governmental policy for Covid-19; suppose that $\sqrt{2}$ is a rational number.

White (1990) characterises supposition as follows:

> To say 'suppose that p' invites or introduces a statement of the consequences or implications of p.

Hence subject S supposes that p iff S finds and states the implications of p. Most authors agree; e.g. (Balcerak Jackson, 2016; Arcangeli, 2019). What S will find out and state, depends on the background knowledge of S. So, I might better put the explicans in the subjunctive mood: S supposes that p iff S finds out and states what p would imply in combination with background knowledge of S if p were true. Now, modal logic enters the party: to say 'if p were true' is to say that p is *possible*. When I further want to emphasise that supposing that p is an occurrent mental state, I arrive at the following explication: S supposes that p iff S occurrently finds and states the implications of $\Diamond p$ in combination with its background knowledge. This, I submit, is equivalent to: S supposes that p iff S occurrently accepts that $\Diamond p$. Which is to say that S is imagining that p (2.13). Indeed, Dokic and Arcangeli (2015, p .11, fn. 17) submit that supposition is "more akin to acceptance rather than to belief"; and Balcerak Jackson (2016, pp. 52–54) argues that supposing must be identified with accepting.

It may be objected that with acceptance I am casting the net too wide: accepting that p encompasses exploring implications of p on the basis of background knowledge, as *supposition* demands, but it also encompasses other mental activities with p. I can correct for this by specifically mentioning that the purpose of supposing that p is exploring the implications of p. Indeed, several authors have emphasised that supposition, in contrast to imagination, *always* has an epistemic purpose; e.g. Balcerak Jackson (2016); Arcangeli (2019); Salis and Frigg (2020). Wanting to know what p implies qualifies as a specific epistemic purpose. So let's put that in.

Then I finally arrive at my explication of supposition:

> [Supposition] Subject S *supposes* that p iff S imagines that p for the epistemic purpose of finding out what p implies. (2.15)

Hence, by virtue of ImProp (2.14), if S supposes that p, then p occurrently accepts that there is a p-world and then S explores which other propositions are then true in that $\langle \tau, p \rangle$-world. If the modality type τ is quotidian or nomic, then S can rely on all scientific background knowledge; if metaphysical, then S can no longer rely on this and must draw implications much more cautiously. Again I note that the choice of τ makes a difference for what the implications of p are (in combination with what we know), which is something that explication (2.15) takes care of.

My proposed explication of supposition (2.15) makes supposition a subtype of proposition-imagination, which is something that has been argued for by fellow Wideheads (Nichols and Stich, 2003; McGinn, 2004). This also harmonises perfectly with the idea that *imagination* is the relevant propositional attitude in *conditional reasoning*, as e.g. (Langland-Hassan, 2020, Ch.5–6), Berto (2022, 2023), Özgün and Schoonen (2022) and Williamson (2016) have explicitly argued.

However, Imagers (2.4) obviously will reject explication (2.15). Balcerak Jackson (2016), an Imager, marks the difference between imagination and supposition as the presence and absence, respectively, of sensory content. The purported difference between imagination and supposition, submitted by Balcerak Jackson (2016, p. 47–48), is that we can suppose anything, that every proposition can be supposed, whereas we can *succeed* or *fail* to imagine a proposition. Think of refusing to, and hence 'failing' to, imagine morally repulsive propositions.[40] But those who point to this difference (like Balcerak Jackson) are Imagers, according to whom the failure to imagine is due to the refusal or resistance to evoke mental states with sensory and affective content of morally repulsive situations.

[40] This is the controversial 'problem of imaginative resistance' (Gendler, 2000a; Gendler and Liao, 2016; Tuna, 2020), c.f. Kim et al. (2019).

Indeed, the request *to suppose* that p will rarely be taken as a request to conjure up imagery or feelings, i.e. it will not be taken to appeal to vivid imagination, whereas *to imagine* that p does generally make an appeal to evoke mental states with such sensory and affective contents. But Wideheads like me, know and acknowledge what is going on here: the refusal or impossibility to imagine concerns only *sub*-types of imagination, what I shall call 'visualisation' (Section 2.5.4) and 'inside-action-imagination' (Section 2.6.1). My in (2.15) explicit mentioning of an epistemic purpose of finding the consequences of p, entails that the supposer is in a 'logical mood', not that the supposer is evoking mental imagery and feelings.

Finally, I emphasise that my explication of supposition (2.15) does *not* mention the actual truth or falsehood of p, *nor* does it mention any belief of S about the actual truth or falsehood of p. In other words, according to my explication (2.15), you can always suppose that p, irrespective of whether or not p is actually true, and irrespective of whether or not you *believe* that p is actually true. Perhaps this strikes one as odd. Can you *suppose* that the Nazi's lost the Second World War, even though you already know this to be true? Balcerak Jackson (2016) thinks so, as she argued that we can always suppose everything (see above), and I think so too. The request to suppose here still achieves something: it is, at the very least, a request to adopt a 'logical mood'.

There is a concept closely related to supposition that *does* denote a mental state that cannot be had when one believes that the supposed proposition is true: counterfactual thought. To this I turn next.

2.5.2 Counterfactual thought

Counterfactual thought is the relevant propositional attitude in counterfactual reasoning, which is the type of reasoning that involves evaluating conditional statements of the logical form: "if p were the case, then q would be the case", or "$p \mathrel{\square\!\!\rightarrow} q$", where p and q denote possible but non-actual states of affairs.

When we counterfactually think that p, what is our mental state?

Just like for supposition, there is widespread consensus that counterfactual thoughts and counterfactual reasoning imply mental states of imagination: when we aim to evaluate a counterfactual conditional of the form "$p \boxminus\!\!\rightarrow q$", we *imagine* some scenario where p is true (for some appropriate modality type), to find out whether q would also be true *in that same imagined scenario*; Byrne (2005, 2016, 2017); Van Hoeck et al. (2015); Epstude and Roese (2008); Epstude (2018); Salis and Frigg (2020). Going even further, Iranzo-Ribera (2022) proposed a full-fledged semantic theory of counterfactuals with imagination at its core; c.f. Kimpton-Nye (2020).[41] In the oft-quoted words of Williamson (2005, p.19):

> When we work out what would have happened if such-and-such had been the case, we frequently cannot do it without imagining such-and-such to be the case and letting things run.

The question that now presents itself, is whether counterfactually thinking that p is at all distinct from *supposing* that p (2.15), which I explicated as imagining that p for the epistemic purpose of finding out what p implies. Yes, it is: counterfactuals are — the name says it all — *counter*factuals; they concern states of affairs that are considered contrary to the facts. In contemporary cognitive science, the concept of counterfactual thought denotes a *psychological* phenomenon, i.e. it denotes the *attitude* that we adopt towards the proposition: namely, that we *believe* that it is false; see e.g. (Epstude and Roese, 2008; Epstude, 2018; Van Hoeck et al., 2015). In other words, we can counterfactually think that p even when p is, unbeknownst to us, also actually true. This gives:

[Counterfactual Thought] Subject S *counterfactually thinks* that p iff S supposes that p and S believes that p is false. (2.16)

This makes counterfactual thought a type of supposition — we should really call it *counterfactual supposition* — that additionally implies a *dis*belief in the supposed proposition.

[41] Even Stalnaker (1968) discussed the role of imagination in counterfactual reasoning, albeit only briefly.

Perhaps one wishes to object to explication (2.16) that the epistemic purpose of counterfactual thought is more restricted than the epistemic aim of supposition: it is not merely to find out what p implies (which, generally, is an infinitude of propositions), it is to find out specifically *whether a specific proposition q would follow from p*, i.e. to evaluate a specific counterfactual *conditional*. This gives:

> Subject S *counterfactually thinks* that p iff there is a proposition q such that S supposes that p for the specific purpose of finding out whether q follows, and S believes that p is false. (2.17)

This explication, which is a special case of explication (2.16), makes sense too. Both Stalnaker (1968) and Lewis (1986), the 'founding fathers' of the now ubiquitous possible-world semantics for counterfactual conditionals, stress that evaluating the truth of a counterfactual conditional is a highly nuanced affair that involves carefully and *selectively* picking out the set of possible worlds where p is true that is *closest to the real world*.[42] If we are not selective in *which* possible worlds we pick out, i.e. if we consider only possible p-worlds that are unrecognizably different from the real world, then we will be unable to 'find' and formulate relevant *evidence* for the truth or falsehood of the counterfactual conditional. All this is vindicated by my explication (2.17): when we counterfactually think that p, we suppose that p in a *specific*, reality-oriented way that makes it possible for us to evaluate whether q follows. We pick out and imagine, in other words, the *appropriate* set of $\langle \tau, p \rangle$-worlds to imagine and reason with. I stress that this "appropriateness" goes beyond merely picking out the

[42] Salis and Frigg (2020) argue explicitly that counterfactual *reasoning* exhibits a certain 'selectivity' and is always 'reality-oriented', and they use this to argue that counterfactual reasoning is a type of propositional imagination that is distinct from supposition, which does not exhibit these characteristics necessarily. I acknowledge these characteristics, as is evident in the main text, but I largely lay aside Salis and Frigg's analysis because they make a category error by intending to discuss types of propositional imagination and then proceed to discuss "counterfactual reasoning" as one of these types. But counterfactual reasoning is a form of *reasoning*, i.e. a mental *process*, not a mental state. The correct concept to analyse is counterfactual *thought*, which *is* a mental state.

appropriate modality τ (per my explication for proposition-imagination (2.14)): it is also to select, within this modality, the *appropriate set of possible worlds* that are closest to the real world as possible, all of which are part of τ.

But I prefer explication (2.16) to explication (2.17). I shall raise two questions to argue for my position. *First*: does having a counterfactual thought *imply* that one will evaluate counterfactual conditionals? In practice, it will generally be the case that a counterfactual thought is followed by evaluating conditionals. But is it conceptually *necessary*? I do not think so. While it is true that counterfactual thought is closely tied to conditional reasoning, it seems too strong to require evaluating a conditional as a *necessary* condition for the mental state of counterfactual thought. So, explication (2.17) is arguably too strong. *Second*: if counterfactual thought does not imply evaluating counterfactual conditionals, then it also does not imply selectively picking out the appropriate set of possible worlds that make it possible to evaluate said conditionals. To illustrate: I can counterfactually think that the Nazi's won the Second World War and, although my mind is 'drawn' to the plausible consequences of this counterfactual thought — I am already in a suppositional 'logical mood' — there is no specific conditional that I am aiming to evaluate. Hence, it seems natural to simply reduce counterfactual thought to supposition, with the additional condition that counterfactual thought implies a disbelief in the supposed proposition, per explication (2.16).

Next up: conceiving.

2.5.3 Conceiving

Conceiving is the relevant propositional attitude in many philosophical thought experiments: Chalmers (1997, 2002) conceives that there exist philosophical zombies; Searle (1980, 1982) conceives that he can communicate in Chinese without understanding Chinese; Putnam (1973) conceives that on twin-Earth, water is not H_2O; Nagel (1980, 2012) cannot conceive

what it is like to be a bat.[43]

One question that immediately obtrudes after the last two Sections: is conceiving distinct from (counterfactual) supposing or are they identical twins? Consider requesting Saskia: *Conceive* that Mary sees red for the first time in her life, but *do not suppose* it. Saskia surely will be at a loss what to do. Next, you request: *Suppose* that the Mary sees red for the first time in her life but *do not conceive* it. Saskia will now be suspecting that you are pulling her leg. Right she is. Either you are able to instruct Saskia what to do when asked to conceive that p and what to do differently when asked to suppose that p (in which case these siblings are phenomenologically distinct), or you must admit they are the same propositional attitudes (in which case they are phenomenologically identical twins), and the different use of the verbs to conceive and to suppose, if so, resides in their specific epistemic purposes. For supposing, the epistemic purpose has been specified, and for conceiving, I shall leave it open, as will transpire below.

Balcerak Jackson (2016, p.56) understands conceiving in terms of *imagined rational belief*:

> When one conceives of p, one [engages] in an exercise of perspective-taking. But one does not take the perspective of the subject as the subject of phenomenal experiences, but rather as the subject of rational belief.

The subject of 'phenomenal experiences' is the imaginer; the conceiver is the imagined rational believer. This gives the following explication:

[Balcerak Jackson] S conceives that p iff S imagines that S rationally believes that p. (2.18)

[43] It was objected by an anonymous referee that it is not evident that *conceivability* is the relevant propositional attitude in these thought experiments. I respond that a brief review of the original articles wherein these thought-experiments were presented show that these authors explicitly *mention* conceiving. Moreover, I have also cited recent articles of Chalmers (2002) and Nagel (2012) where these authors *explicitly* discuss conceivability and its relation to possibility. I take this to be conclusive evidence that, at the very least, it was these authors' *intention* that conceiving is the relevant propositional attitude in the thought experiments.

Yablo (1993, p. 22) ascribes a similar view to Putnam:

> To conceive a proposition, in Putnam's sense, is to imagine acquiring evidence that justifies you in believing it.

If Balcerak Jackson's "rational belief" and Putnam's "acquiring evidence that justifies" believing it are the same, then they share their conception of conceiving.

Yablo (1993, p. 22) begs to disagree with conception (2.18): we can conceive, for example, that we were never born, or that we do not have any beliefs, whereas we cannot imagine rationally believing *that* on pain of ending up in contradiction. Yablo (*ibid.*) elaborates delicately:

> I find p conceivable if I can imagine, not a situation *in* which I truly believe that p, but one *of* which I truly believe that p.

About "truly" (*ibid.*, p. 23):

> To believe truly is to believe a truth, so you imagine a situation in which you believe some true proposition.

Then, in terms of possible worlds, Yablo says: if S conceives that p, then S imagines a possible world W and *truly* believes that W is a p-world, so that, then, $\Diamond p$ is true *in our world* due to the "truly". This gives:

[Yablo1] Subject S conceives that p iff S imagines some world W such that S believes that p is true in W, and p is true in W, so that $\Diamond p$ is true (in our world). (2.19)

Elsewhere in the same paper, after having considered several conceptions of conceiving set forth by philosophers in the past, Yablo (1993, p. 29, fn. 59) writes:

> (a) 'I conceive that p' iff I imagine a world which I take to verify p;
> (b) 'p is conceivable for me' iff I *can* conceive that p.

When we take the phrase takes to as belief, this yields the following explication of conceiving:

> [Yablo2] Subject S conceives that p iff S imagines a world W, such that S rationally believes that p is true in W, so that S also believes that $\Diamond p$ is true (in our world). (2.20)

Explication [Yablo2] (2.20) is logically weaker than [Yablo1] (2.19): it dispenses the conjunct that p *is* possible. Which explication is preferable, must be decided on the basis of the cardinal issue whether 'conceivability is a guide to possibility'. The question whether conceivability implies possibility is trivially decided on the basis of [Yablo1] (2.19): if any S manages to conceive that p, then p is possible. Explication [Yablo2] (2.20) fails to deliver this result, because S believing that $\Diamond p$ does not imply that $\Diamond p$ is true.

I join Yablo in rejecting [Balcerak Jackson] (2.18). But I am tempted to propose a much simpler explication of conceiving than Yablo's (2.19) and (2.20), which nonetheless is similarly illuminating, as follows:

> [Conceiving] Subject S conceives that p iff S imagines that $\Diamond p$ for some epistemic purpose. (2.21)

Substituting in (2.21) my explication of ImProp (2.14), we obtain:

> Subject S conceives that p iff S occurrently accepts that it is τ-*possible* that $\Diamond p$, for some appropriate modality type τ, for some epistemic purpose. (2.22)

The concept of possibility now occurs twice in succession. This is convenient: when the two modality types the same (and I have no reason to assume otherwise), conceiving that $\Diamond p$ collapses to conceiving that p, due

to the following theorem of modal logic:

$$\vdash \Diamond\Diamond p \longleftrightarrow \Diamond p\,. \tag{2.23}$$

I thus end up with the following Wideheadian biconditionals:

> S conceives that p iff S imagines that $\Diamond p$ for some epistemic purpose iff S imagines that p for some epistemic purpose. (2.24)

The obtruding question of whether there is a difference between supposing and conceiving can now be answered: supposing is a sub-type of conceiving, where the epistemic purpose is finding out what p implies (2.15). And concerning the cardinal issue whether conceivability implies possibility, I answer in the negative, because accepting that $\Diamond p$ does not imply that $\Diamond p$.

So much for supposition, counterfactual thought and conceiving. I move on to more 'sensorial' concepts: visualisation and picturing.

2.5.4 Visualisation

A visualisation is always a visualisation *of* something: you can visualise "London or the fictional city Macondo, Napoleon or Raskolnikov, a tiger or a unicorn" (Salis and Frigg, 2020, p. 27). Visualising a proposition presupposes that the proposition is *visualisable*, which raises the question how encompassing the class of all visualisable propositions and entities is. I must first inquire into this issue to determine the *scope* of visualisation.

Tempting is to consider visualisability being the same as observability. After all, both are intimately connected to visual perception. Concrete observable objects can certainly be visualised, and, since philosophers of science have explicated observability, I can refer to their explications and be done with it.[44] *Alas!* Fallen for the wrong temptation.

One reason why this does not work is that *unobservable* concrete objects can be visualised too. We can visualise a miniaturised human being inside a submarine, propagating in the blood stream of an ordinary human

[44] van Fraassen (1980), Muller (2005).

Figure 2.5: Visualisation of a closed and an open interval of real numbers: $[-1, +1] \subset \mathbb{R}$ and $(-1, +1) \subset \mathbb{R}$, respectively, by line segments, dots and circles.

being, avoiding collisions with white and red blood-cells, viruses, bacteria and other *un*observable entities. The old motion picture *Fantastic Voyage* (1966) visualises this in all splendour.

Another reason is that *abstract* entities can be visualised too. Mathematicians draw diagrams and pictures on a regular basis to visualise abstract objects: an interval of real numbers can be visualised by a straight line segment (Figure 2.5); a function can be visualised by an arrow diagram; intersections, unions and complements of sets are visualised by Euler-Venn diagrams; Feynman diagrams visualise terms in perturbation series; and so on and so forth.

Mathematical entities having a high level of abstraction are candidates for *unvisualisable* entities. But suppose Sonja draws an ellipse and shades its interior, draws a curve with an arrow-head from rectangle to itself passing through the ellipse, and asserts that the ellipse is a Von Neumann ring and the rectangle is Hilbert-space (see Figure 2.6). Then her drawing qualifies as a visualisation. Sonja draws Eilenberg-MacLane diagrams that visualise relations in category theory, the most abstract realm in mathematical discourse. Visualisations can have almost nothing in common with what they visualise — a visualisation of something is not a *picture* of it.

These two reasons suggest that *any* entity can be visualised.

Furthermore, unlike the ability to see, which depends on our sensory organs and is therefore more-or-less the same for all human beings with healthy eyes and brains, the cogitative capacity to visualise is not evenly spread among all human beings, but will depend heavily on the individual

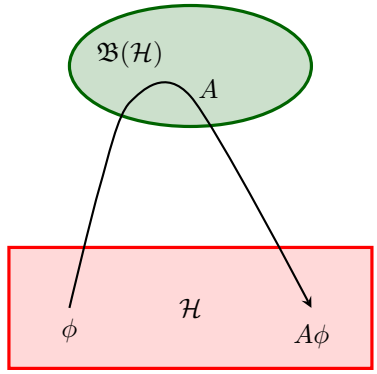

Figure 2.6: *Visualisation.* Bounded operator A from Von Neumann ring $\mathfrak{B}(\mathcal{H})$ sends Hilbert-vector $\phi \in \mathcal{H}$ to Hilbert-vector $A\phi \in \mathcal{H}$.

human being under consideration. Choices must be made with respect to *how* we aim to construe visualisability. If, for example, (i) an entity is visualisable iff it can be visualised by *every* human being, then the class of visualisable entities will become a tiny proper subclass of the class of all entities, I fear. If, by contrast, (ii) an entity is visualisable iff it can be visualised by *some* human being, then the visualisable entities will coincide with the class of all entities. Which should we choose? I prefer (ii), if only to prevent that the concept of visualisability will be taken hostage by *aphantasiasts*, who self-report being unable to visualise *anything*; see Section 2.8 for discussion of *aphantasia*.

I conclude that I have every reason to hold that *all* entities are visualisable: the class of visualisable entities coincides with the class of all entities. By this same line of thought, I arrive at the visualisablility of all propositions: the class of visualisable propositions coincides with the class of all propositions. (Even contradictory propositions can be visualised; recall Escher's impossible cube, Figure 2.4.)

I next turn to explicating proposition-visualisation.

Both visual perceptions (2.1) and visualisations are mental states with visual content and intentional objects. The difference is that visual per-

ceptions are exogenous (and typically veridical) while visualisations are endogenous (and typically falsidical). This leads to the following putative conceptual truth: if S visualises that p, then S has an occurrent endogenous mental state with visual content having the topic of p as its intentional object. Is this necessary condition also sufficient?

Nope. Let's see why.

Wideheads Salis and Frigg (2020, p. 29) write that imagination:

> ... cannot be defined in terms of the presence of mental images because mental images can accompany episodes of memory, belief, desire, hallucination, and more. What makes the deployment of a mental image an instance of imagination is the *attitude* we take toward the mental image.

Having an endogenous mental state with visual content is not sufficient for imagination, because such a state can be also any of the other mental states listed by Salis and Frigg above.

I have already distinguished imagination from hallucinations and from beliefs (Section 2.4.2). Concerning desires, quick and dirty: desires are mental states with affective content and worldly satisfaction conditions, unlike imaginings; and they have, as Anscombe called it, a direction of fit from world to mind, whereas imaginings do not have such fitting relations. I shall distinguish imagination from memory in Section 2.6.3. I turn now to visualisation.

I have already identified the distinguishing *attitude* of ImProp: occurrent *acceptance*. For the case of visualing a proposition, we need clarity about the visualisation-of relation. *Representation* is the obvious choice. Then visualisation becomes, in a nutshell, visual representation: the visual content *represents* p. I need not commit to a specific understanding of *representation* here: you may use your own preferred account. I reemphasise that, if one's preferred account of representation does not allow mental content to represent due to metaphysical issues (e.g. the idea that only concrete objects can represent, c.f. (Salis et al., 2020)), then one in-

stead can pose the condition that S only *accepts* that the visual content represents p (even though S disbelieves this) — I am after the *attitude* that S adopts to the visual content, not the metaphysical relation between the visual content and the world.[45] Thus I propose to explicate proposition-visualisation as follows:

> [Proposition-Visualisation] Subject S *visualises that* p iff S has an occurrent endogenous mental state such that its semantic content is $\Diamond p$, which S accepts, and its visual content represents p. (2.25)

By virtue of explication of ImProp (2.13), we have immediately that if S visualises that p, then S imagines that p (in the context of some appropriate modality type τ). This makes proposition-visualiation a subtype of ImProp, as it should be.

Besides visualising propositions, we can, and often do, visualise entities and actions. Action-visualisation I shall explicate in Section 2.6. Entity-visualisation we can explicate as either (i) visualising *perceiving* the entity, which, by (2.6), is an instance of action-visualisation (Section 2.6); or (ii) *nearly* visualising *that* the entity exists, per the following explication:

> [Entity-Visualisation] Subject S *visualises entity* ε iff S accepts that ε possibly exists, for some appropriate type of modality, and S has an occurrent endogenous mental state such that its visual content represents ε. (2.26)

Entity-visualisation is *nearly*, but not entirely, visualising *that* ε exists (which is proposition-visualisation). The only difference is that the acceptance in my explication of entity-visualisation (2.26) is not occurrent

[45] To motivate this further, I appeal to the important observation from Wyer Jr. (2007, p.285) that "[t]heories of mental representation are inherently metaphorical. They must consequently be evaluated on the basis of their utility and not [only on] their validity in describing the physiology of the brain." See also (Salis et al., 2020) for a recent discussion about (pretend)-representation by mental states.

but dispositional. Only if S were asked what S is doing, and S were to respond I am imagining ε, then, at that moment, the mental state with this semantic content would become occurrent, thus would become an instance of proposition-visualisation (2.25). The reason why the semantic content is not accepted occurrently is to prevent that the occurrent mental state of visualisation would have semantic content, which would incorrectly turn entity-visualisation into an occurrent propositional attitude.

As promised in Section 2.4.2, I next use my explication of proposition-visualising (2.25) to obtain an explication of proposition-*imagination for Imagers*, according to whom mental imagery is necessary for imagination. Replacing in the explicans in (2.25) the visual content with *sensory* content, a.k.a. mental imagery, we obtain:

> [ImProp for Imagers] Subject S imagines that p iff S has an occurrent endogenous mental state such that its semantic content is $\Diamond p$, which S accepts, and its sensory content represents p. (2.27)

The Imagers' Explication (2.27) vindicates the eight features of imagination (I)–(VIII) in almost the same way as (2.14) does. *Almost!* Two important differences. *First*, ImProp for Imagers (2.27) is even *more* Perception-like (VII) than ImProp, as it should be, because it necessarily includes sensory content. *Second*, assuming that, if the sensory mental state is endogenous, then it is not exogenous, then Wittgenstein's remark (quoted on p.51) is vindicated: when you visually perceive (2.1) an object, then you can neither visualise it (2.25) nor, according to Imagers, imagine *that* it exists (2.27). Visualisation (2.25) and ImProp for Imagers (2.27) are *not* independent of visual perception: they are mutually exclusive.

2.5.5 Picturing

John Lennon wrote and sang with The Beatles in *Lucy in the Sky with Diamonds* (1967):

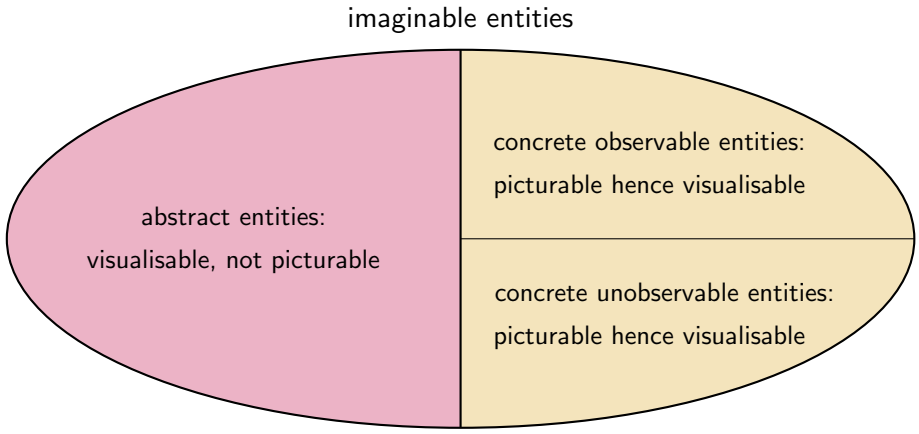

Figure 2.7: Different types of entities and how they can be imagined.

Picture yourself on a boat on a river
With tangerine trees, and marmalade skies.
(. . .)
Picture yourself on a train in a station
With plasticine porters with looking glass ties.

Earlier I ascertained that a visualisation of p, notably when the topic of p is unobservable or abstract, need not have much in common with p. Picturing is different. Picturing is a constrained form of visualisation in two ways.

Firstly, just as it makes no sense *to make photographs* of abstract objects, it also makes no sense *to make mental pictures* of abstract objects.[46] Only concrete objects can be pictured: an entity can be pictured iff it is concrete (see Figure (2.7)). *Secondly*, when asked to picture that p, not any old random visual mental imagery will do, that is, the visual content of the mental state needs to reach some level of accuracy for it to qualify as a *picture* of p.

[46] It is a conceptual impossibility to make photographs of abstract object, but this doesn't mean that we don't *try*; Costello (2018).

I propose the following explication of picturing:

> [Proposition-Picturing] Subject S *pictures that* p iff S has an occurrent endogenous mental state such that its semantic content is $\Diamond p$, which S accepts, and its visual content accurately represents p. (2.28)

In a nutshell: picturing is accurate visualising.

I note that the distinction between visualisation (2.25) and picturing (2.28) is not sharp but *vague*, because the difference between inaccurate and accurate representation is *gradual*, and when the accuracy is low (only few properties and relations of the target correlate to those of the visual content), picturing morphs into 'mere' visualisation, which seems appropriate.[47]

Entirely analogous to visualising an entity (2.26), we have picturing an entity:

> [Entity-Picturing] Subject S *pictures* entity ε iff S accepts that ε possibly exists, for some appropriate type of modality, and S has an endogenous occurrent mental state such that its visual content accurately represents ε. (2.29)

So much for picturing and visualising. Five concepts allied to proposition-imagination have now been explicated. I next turn to explicating my final type of imagination: action-imagination (ImAct).

2.6 Action-imagination

2.6.1 Explicating action-imagination

In Section 2.3, I briefly discussed and adopted a widely shared distinction between two types of imagination: *proposition*-imagination (ImProp),

[47] Question: is the book-title *Picturing Quantum Processes* by Coecke and Kissinger (2017) appropriate; can we *picture* quantum processes or can we only *visualise* them?

which I explicated as occurrent acceptance of a possibility of an appropriate modality type (2.14), and *action*-imagination (ImAct), which I shall explicate next. Since the considerations that have led me to my previous explications — notably, ImProp (2.14) and visualisation (2.25) — are similar to the considerations that motivate my explication of ImAct, I shall proceed in an accelerated fashion whenever possible.

I begin by drawing an important distinction between (i) imagining performing an action from a 1st-person perspective, and (ii) imagining performing an action from a 3rd-person perspective. If Joel imagines swimming in the Dead Sea from a 1st-person perspective, he imagines not only what the Dead Sea *looks, smells, sounds* and even what the extremely salty water *tastes like*, he also imagines what swimming in the Dead Sea *feels like*, experiencing the sense of his buoyant body — his *own* body — floating nearly on top of the surface. If, by contrast, Joel merely imagines swimming in the Dead Sea from a 3rd-person perspective, he adopts what Walton (1990, §1.4) calls a "spectator perspective", *seeing himself* swimming in the Dead Sea, perhaps from the beach or from a bird's-eye view, without the associated sensations of buoyancy and feelings of self-involvement. (See a spectator perspective in Figure 2.8.) Likewise, Dokic and Arcangeli (2015), following e.g. Vendler (1984) and Peacocke (1985), distinguish imagining action 'from the inside' and imagining action 'from the outside'.

The difference between these two types of action imagination is phenomenological: inside-action-imagination is accompanied by an explicit sense of embodiment, agency and self-involvement — a sense of being immersed in the activity, *as if* actually doing it yourself — whereas outside-action-imagination is not. The psychologist Jeannerod (1994) famously distinguished between 1st-person and 3rd-person mental imagery: 3rd-person mental imagery is *sensory* content, i.e. the type of imagined content that has its counterpart in perception, while 1st-person mental imagery is *motor* content, i.e. a type of imagined content that has its counterpart in our "motor physiology" (*ibid.*, p. 189). Following this tradition, I shall

Figure 2.8: My friend Joel "swimming" in the Dead Sea.

describe the explicit 'sense' of embodiment, agency and self-involvement as a mental state with endogenous *motor* content. I emphasise that the notion of 'endogenous motor content' is subtle, because motor content is *always* internally caused. With *endogenous motor content*, I mean motor content whose causes *and* effects are internal: endogenous motor content is "the representation that results from processing in the motor system (in the motor and premotor cortices) *that does not trigger motor output directly*" (Nanay, 2021, §4.1, my italics). Endogenous motor content does not *directly* correspond to physical action (even though it can indirectly *motivate* physical action; recall Section 2.4.2, feature V. *Quarantined*).

All things considered, I propose the following explications for inside-action-imagination (ImActIn) and outside-action-imagination (ImActOut), based on the distinction that ImActOut requires only sensory content, and

ImActIn requires only motor content:

> [ImActIn] Subject S *inside-imagines* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate type of modality, and S has an occurrent endogenous mental state such that its sensory content represents the event of S ϕ-ing.
>
> [ImActOut] Subject S *outside-imagines* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate type of modality, and S has an occurrent endogenous mental state such that its motor content represents the event of S ϕ-ing. (2.30)

The explication for action-imagination *in toto*, then, is the inclusive disjunction of ImActIn and ImActOut (2.30):

> [ImAct] Subject S *imagines* ϕ*-ing* iff S inside-imagines ϕ-ing or S outside-imagines ϕ-ing, or both. (2.31)

I emphasise that ImActOut and ImActIn do not exclude each other and, in fact, often (but not necessarily) come hand-in-hand in practice; see Section 2.8. I next make two systematic remarks about explications (2.30).

First, in explication (2.30), the condition that S *accepts* that S is ϕ-ing is motivated by the thought that if S does not accept that it is possible for S to ϕ, then S does not *imagine* ϕ-ing. If you imagine flying like Superman, then you accept that it is metaphysically possible to fly like Superman. I stress that S is not accepting the possibility *occurrently*, for otherwise the occurrent mental state of imagining ϕ-ing would have had also semantic content, thus making it an instance of ImProp (2.14). But this would be incorrect because ImAct is not ImProp: the two are distinct types of imagination.

Second, explication (2.31) also accounts for all features of imagination (I)–(VIII) of Section 2.4.2, as one may care to verify. Note that I did *not* pose the condition that S *is not* ϕ*-ing* as a conjunct to the explicans. If I had done so, then imagining ϕ-ing and 'actually' ϕ-ing would

have excluded each other, and action-imagination would no longer be Independent (VI). I suggest to leave open the possibility that S is ϕ-ing and simultaneously imagining ϕ-ing. Think, for example, of drawing an elephant with your eyes closed, whilst simultaneously visualising yourself drawing an elephant. Yet anyone who subscribes to the mentioned exclusion (like Wittgenstein, recall discussion on p.51), may add the mentioned conjunct.

2.6.2 Visualisation and picturing revisited

In the previous Section, I explicated the inside and outside types of ImAct. Analogously to my explication of outside ImAct (2.30), I can explicate *action*-visualisation as follows:

> [Action-Visualising] Subject S *visualises* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate modality type, and S has an occurrent endogenous mental state such that its visual content represents the event of S ϕ-ing. (2.32)

Mutatis mutandis for action-picturing:

> [Action-Picturing] Subject S *pictures* ϕ-ing iff S accurately visualises ϕ-ing. (2.33)

When the acceptance in (2.32) and (2.33) becomes occurrent, then S visualises or pictures the proposition *that* S is ϕ-ing, per (2.25) and (2.28), respectively, which is exactly what I want.

Finally, recall my reduction of sensory entity imagination (2.6) to action-imagination: S sensorily imagines entity ε iff S imagines perceiving ε. When we now substitute the explication of action-visualisation (2.32)

in the right-hand side of (2.6), we obtain:

> S visually imagines entity ε iff S accepts that it is possible that S perceives ε, for some appropriate modality type, and S has an occurrent endogenous mental state such that its visual content represents the event of S seeing ε. (2.34)

Note that, if it is possible that S sees ε, then there is a possible world where S sees ε. In that world ε exists, so if S accepts that it is possible that S sees ε, then S also accepts that ε possibly exists. If S is seeing ε, then by my explication of vision (2.1), the visual content of the mental state visually represents ε accurately. I now arrived at the explicans of [Entity-Picturing] (2.29). Hence, if S imagines seeing ε, then S pictures ε; and conversely. My putative explication of S visually imagining concrete entity ε as S imagining seeing ε is in perfect logical harmony with my explication of S picturing ε. The logical cement in my conceptual reticulum hardens.

2.6.3 Memory

The time has come to distinguish imagination and memory. Recall that Salis and Frigg (2020, p. 29) wrote the following:

> [Imagination] cannot be defined in terms of the presence of mental images because mental images can accompany episodes of memory, belief, desire, hallucination, and more. What makes the deployment of a mental image an instance of imagination [rather than e.g. memory] is the *attitude* we take toward the mental image.

If Salis and Frigg hold that mnemonic and imaginative *content* are, or ought to be, always mutually exclusive, then I disagree. If Salis and Frigg hold, *contra* Hume and other empiricists from the past, that these contents cannot be discerned *phenomenologically*, then I agree.

To begin, I wish to argue that memory and imagination are not mutually exclusive. Consider the following example. On request, Samantha

remembers the playground at her primary school, which results in a mental state with mnemonic content. Next, Samantha is asked to imagine seeing a pink elephant on that same playground. This results in a mental state of memory *with* imagined content — or, equally true: a mental state of imagination with mnemonic content. I conclude that mental states of imagination can have mnemonic content: memory and imagination are not mutually exclusive, in contradistinction to what Salis and Frigg suggest.

Since Martin and Deutscher (1966), see also Fernandez (2008); Bernecker (2009, 2017); De Brigard (2014a,b, 2020), c.f. Michaelian (2016a,b); McCarroll et al. (2022), there has been reasonable consensus on the idea that memory entails the existence of an 'appropriate causal connection' between (the subject's experience of) a past event and the content of a mental state of memory. If Samantha remembers ϕ-ing, then she ϕ-ed in the past, and her past ϕ-ing causally influenced her current mental state in one way or another; if Samantha 'merely' imagines ϕ-ing, then she need not have ϕ-ed in the past at all.[48] Thus, I can explicate mnemonic ImAct as follows:

> [ImActMem] Subject S *mnemonically imagines* ϕ-ing iff S imagines ϕ-ing, S ϕ-ed in the past, and there is an appropriate causal connection from S's ϕ-ing in the past to the sensory and motory content of S's current mental state of imagination. (2.35)

I take (2.35) to be essentially correct. However, explication (2.35) does not give us an *attitude* that distinguishes imagination from memory, which is what Salis and Frigg were after (see quote above). Well, Russell (1921, p. 186) already identified this attitude: "Memory demands (a) an

[48] I wish to flag that we must handle the notion of "appropriate causal connection" with care. In light of Hume's *recombination principle* (Section 1.2), it can be plausibly argued that *all* imagined content is also based on, i.e. *has a causal connection to*, past experiences, in one way or another. So, the phrase "appropriate" in 'appropriate causal connection' serves to distinguish the causal connections that determine mnemonic content from the (arguably less *direct*) causal connections that determine imagined content. I lay aside the question whether there is a principled way of working this out.

image, and (b) a belief in the past existence." Urmson (1967) famously argued for this same point: memory entails a *belief* that the remembered event happened in the past — which, in principle, can be checked from a third-person perspective, i.e. by someone else than the rememberer.[49] This would give:

> Subject S *mnemonically imagines* ϕ-ing iff S imagines ϕ-ing, S truly believes that S ϕ-ed in the past, and there is an appropriate causal connection from S's ϕ-ing in the past to the content of S's current mental state. (2.36)

It is true that memory is often accompanied by a belief about the past. But I am unconvinced that memory *necessarily* entails such a belief. Counter-examples have been discussed in the literature. I review the most common counter-example: the case of 'accidental' remembering.

Think of a painter who paints a landscape at old age and who is *convinced* that she just imagined this landscape, rather than remembered it, but she does not realise that the landscape she just painted strongly resembles the landscape surrounding her childhood home — this childhood landscape is causally responsible for the painted landscape (Liao and Gendler, 2020). Intuitively, we would want to characterise this landscape as a *remembered* landscape: we want to say that, when the painter visualised the landscape that she was about to paint, she *remembered* the landscape rather than imagined it. But the painter is convinced that she did not remember: she does not believe that she saw this landscape in the past. This makes the second condition in (2.36) is false, and, consequently, (2.36) wrongly tells us that the painter is not remembering while she actually is. Explication (2.35), however, *does* correctly denote the painter's mental state as a state of memory.

I conclude that memory does not *necessarily* imply an attitude —

[49] Along similar lines, Fernandez (2008) argued that memory entails a belief about the *appropriate causal connection* from the past event to the current mental state of memory (2.35). But this seems to strong: someone who does not believe in causality can remember too.

a belief about the past — that distinguishes it from imagination, even though memory may *often* be distinguishable in practice by a belief about the past.

I have argued that mental states of imagination can have mnemonic content, and I explicated mnemonic imagination (2.31). Mnemonic imagination necessarily entails the existence of an 'appropriate causal link' between an (mnemonically) imagined event and the actual occurrence of that event. Moreover, while there is not *always* a phenomenological difference between memories and imaginings, memory *often* entails a belief: the belief that the imagined event happened in the past. This belief about the past is the attitude that Salis and Frigg (2020) were after.

Let me now consider the relation between memory and imagination from the other side: are mental states of imagination *always* mental states with mnemonic content, or vice versa? The answer to this question depends on the explications of memory and mnemonic content, to which I turn now. I begin with a few remarks on the relation between memory and imagination.

Memory and imagination have long been regarded as importantly distinct in many ways: conceptually, cognitively, epistemically, etc. In recent years, however, the idea that memory and imagination are 'continuous' with each other rather than fundamentally distinct — that both are exercises of the *same* mental faculty — is rapidly gaining popularity both in philosophical and scientific debates; c.f. Michaelian (2016a,b); Michaelian and Sutton (2017); Hopkins (2018); (Liao and Gendler, 2020, §2.4). This *continuity hypothesis* is made especially plausible by recent empirical research that demonstrated that our ability to remember the past and our ability to imagine the future are mostly grounded in the *same* neural mechanisms; see e.g. Squire et al. (2010); Schacter et al. (2011); De Brigard (2017); Sant'Anna et al. (2020); c.f. Munro (2021).

I note moreover that the most prevalent distinction between different types of imagination — the distinction between proposition-imagination and action-imagination — has a direct counterpart in memory: within the

concept of memory, too, the most prevalent distinction is between *proposition* (or 'semantic' or 'factual') memory and *episodic* (or 'recollective' memory); c.f. (Michaelian and Sutton, 2017, §2.1). Proposition memories have semantic content: they are concerned with facts, often with facts about "the world in general" (*ibid.*, §2.1.1); e.g. you remember *that* Asian elephants have smaller ears than African elephants. Episodic memories, by contrast, are recollections of past *perceptions*;[50] e.g. you remember *seeing* an African elephant. Like the content of action-imagination (2.31), the content of episodic memories is sensory and motory, rather than semantic. Moreover, like the content of action-imagination (and unlike the content of perception), the content of episodic memory is *endogenous*, rather than exogenous. Pointing to the similarities between these two parallel distinctions in memory and imagination is not a decisive argument in favor of the idea that that memory and imagination are logically connected, but it surely adds further plausibility to it.[51]

Proceeding with the idea that, per the 'continuity hypothesis', memory and imagination are closely connected rather than fundamentally distinct, I next propose explications of episodic memory and propositional memory, which turn out to exhibit a remarkable difference with respect to their relation to imagination.

Episodic memory

On the basis of the above-mentioned considerations, it is tempting to suggest an explication of episodic memory *as* action-imagination directed at the past, in conjunction with the added conditions that the remembered

[50] More generally, episodic memories are recollections of past *experiences*. I however avoid the use of the ambiguous concept of experience in this Thesis, thus limit the content of episodic memories to recollections of (at least) past perceptions. [51] Additionally, I note that in contemporary literature there is widespread agreement that memory is *psychologically generative*: "memory processes do not simply retain the experienced content from the time of the original experience, but actively manipulate them in transformative ways" (Miyazono and Tooming, 2023a, p.129). This casts further doubt on the idea that there is a sharp distinction between memory and imagination. *Note:* I have reduced this important comment to a footnote because in this Thesis I proceed with a simplified understanding of memory as "psychologically preservative".

past event actually happened and is *appropriately* causally responsible for the content of rememberer's current mental state. This was recently explicitly argued for by e.g. Hopkins (2018): episodic remembering is imagination controlled by the past. But this is just my explication of mnemonic imagination (2.35). Thus:

> [Episodic memory] Subject S *remembers* ϕ-ing iff S mnemonically imagines ϕ-ing (2.35). (2.37)

Since this explication does not involve a belief, it stands tall in the face of the counterexample mentioned above. Per explication (2.37), episodic memory allows for 'accidental' remembering: one can remember an event *without* believing that they are remembering, and even while *disbelieving* that they are remembering.

I emphasise that episodic memory (2.37) is *factive*, in the sense that one can remember an event only if they actually perceived it in the past — hence only if the event actually happened in the past. On this there is reasonable consensus in the literature; recall the references on the 'appropriate causal connection' above, see also (Bernecker, 2009, §1.5).

There is however an ostensible conceptual tension between the factivity of memory and the common usage of the phrase *incorrect*, *wrong*, or *false* memory. It is often said that we can remember something *wrongly*; that we can remember something that did not actually happen. This way of using the word "memory" is ubiquitous but it is also misleading: we cannot falsely remember, we can only *falsely believe that we remember*. (Typically, but not necessarily, if we falsely believe that we remember an event, then we *imagine* the event. This is reflected in the colloquial phrase: "am I remembering it or am I merely imagining it?"). Of course, the fact that episodic memory is factive does not imply that beliefs gained on the basis of what we take to be memories are *infallible*. Memory is factive, but beliefs *about* memory are not: if we falsely believe that we remember, then we can obtain a false belief about the past on the basis of what we

believe to be a memory (but which actually is not a memory).

What may add to the confusion about the (misleading) notion of 'incorrect memory' is that there *is* such a thing as correct memory, which is the case when we remember and we *truly believe* that we remember. Notwithstanding, given the fact that incorrect memory is a misleading term, I suggest we avoid use of 'correct or incorrect memory' altogether. We should instead use 'correctly or incorrectly *believing that we remember*', which is devoid of ambiguity.

I note, finally, that understanding episodic memory as imagination directed at and constrained by the past, per (2.37), sheds yet another interesting light on the *voluntariness* of imagination (hence memory): generally, the *moment* of remembering can be voluntarily chosen, and the *topic* of memory can also be voluntarily chosen — i.e. we may freely choose to remember, say, either ϕ_1 or ϕ_2 — but neither the *appropriate modality* of the accepted possibility nor the the *episodic content* of memories can be voluntarily chosen, at least not 'as voluntarily' as we can choose the content of 'pure' imagination, for the simple reason that the content of our memory (if correct) is determined by the remembered past event through an appropriate causal connection.

This relates to so-called 'non-believed memories', which occur when subjects have what they take to be "vivid autobiographical memories for an event but stop believing that the event occurred" (Li et al., 2020, p.1277). I note that the phrase "non-believed memory" is misleading in one of two ways. If the mental imagery that the subject takes to be a memory is *involuntary*, then the subject (i) hallucinates ϕ-ing (*à la* Dostoevsky's polar bear, recall Section 2.4.2), and (ii) falsely believes that they are remembering ϕ-ing, *and* (iii) truly believes that they did not ϕ in the past. If the mental imagery is *voluntary* instead, then the subject (i) imagines ϕ-ing, rather than hallucinates ϕ-ing, and conditions (ii) and (iii) remain the same. S is remembering in neither case: the phrase "non-believed memories" is a misnomer — they are *hallucinations or imaginings mistaken for memories*, coupled, rather jarringly but logically consistently,

with the true belief that the event never happened.

Propositional memory

In contradistiction to episodic memory, which, I argued, does *not* necessarily involve beliefs, most authors understand propositional memory to *be* the recollection of past knowledge — hence the recollection of past beliefs. If you did not believe that p in the past, then you cannot remember that p in the present.[52]

Malcolm (1963, p.236), for example, explicated propositional memory as follows:[53]

> [Malcolm] S remembers that p iff (i) S knows that p, and (ii) S knew that p in the past, and (iii) if S had not known that p in the past, then S would not know that p now. (2.38)

Malcolm's condition (iii) is meant to secure the 'appropriate causal connection' between the knowledge in the past and the current remembered knowledge. But Malcolm's way of phrasing this condition in this way, i.e. as a counterfactual conditional, makes it vulnerable to counterexamples. I mention one counterexample put forward by Zemach (1968, p.527):

> Suppose I came to know that p (a fact about my family) [...] [in the past, but] [...] I forgot this fact later on. At the present, however, I am examining some old family documents, and, upon encountering an entry saying that p, I suddenly remember that, indeed, p. I have been reminded of this fact. Now, clearly, even if I had not remembered that p I would have now known that p, since I have adequate evidence that p (I found a document saying so). But, as one may put it, something else happened, too. Not only have I learned that p; I have also *remembered* this long-forgotten fact. But on Malcolm's analysis [...] it would be impossible for me to say that now I *remember* that p, because my past knowledge that p is not a

[52] See (Bernecker, 2009, §1.5) for a rare exception. [53] I omitted references to time (t_1, t_2) from Malcolm's explication (2.38) to make the similarity with Zemach's explication (2.40) more apparent (see below).

> *necessary* condition for my present knowledge that p. On this view, apparently, I must not say that I have remembered anything.

Thus, Malcolm's explication is too narrow: it does not identify as memory the above-described example which *should* be identified as memory. Zemach (1968) proposed the following alternative explication:

> [Zemach] S remembers that p iff (i) S believes that p, and (ii) if S believes that p then S knows that p, and (iii) if S knows that p then S knew that p in the past, and (iv) S believes that they knew that p in the past. (2.39)

I make two brief remarks about this explication (2.39).

Firstly, conditions (i)–(iii), which are propositions of the form p and $p \to q$, contain logical redundancies. If we consider the logical equivalence that $\big(p \wedge (p \to q)\big) \longleftrightarrow \big(p \wedge q\big)$, and apply this to conditions (i)–(iii) twice, and we use that fact that knowledge implies belief, then we can simplify Zemach's explication (2.39) to:

> [Zemach simplified] S remembers that p iff (i) S knows that p, and (ii) S knew that p in the past, and (iii) S occurrently believes that S knew that p in the past. (2.40)

I note that I took the liberty to add to Zemach's original condition (iii) the condition that S's belief about the past is *occurrent*, which is necessary because else condition (iii) follows from condition (ii) due to the reflexivity of knowledge. I hold that my addition is justified given that Zemach holds that propositional memory implies a belief about the past — an *occurrent* belief, that is.

Secondly, this version of Zemach's explication (2.39) shows us clearly that the only difference between Malcolm's explication (2.38) and Zemach's (2.39) is condition (iii). Zemach's explication does *not* demand an 'appropriate causal connection' for memory, but it *does* require a belief. As I

have argued above, I side with Malcolm on this matter: I hold memory does *not* necessarily require a belief, but it *does* require an appropriate causal connection between the subject's past knowledge — or, perhaps more accurately, the event of knowledge acquisition — and the subject's current state of memory. Those who disagree with this may choose to adopt Zemach's (simplified) explication (2.40), or they may add Zemach's condition (iii) to my proposed explication for propositional-imagination, presented below.

To safeguard Malcolm's explication (2.38) against Zemach's above-mentioned counterexample, we can change condition (iii) into the condition that there is an appropriate causal connection from the past knowledge to the current, remembered knowledge, just like is the case for *episodic* memory (2.37). This gives:

> [Propositional memory] S remembers that p iff S knows that p, and S knew that p in the past, and there is an appropriate causal connection from S's past knowledge that p to S's occurrent knowledge that p. (2.41)

I could go on forever about the merits and shortcomings of each explication of propositional memory discussed so far. But I will not do so, because there is a much more important point I wish to flag about these explications of propositional memory (2.38), (2.39) and (2.41): *none of these explications of propositional memory involves imagination.*

There is thus a striking contrast between the two types of memory and their relation to imagination: episodic memory (2.37) was explicated *as* imagination directed at the past, but propositional memory does not involve imagination at all. While this may strike us as strange, this difference between these two types of memory and their relation to imagination is easily explained. Episodic memory is a mental state with endogenous sensory or motory content: it is a mental state with *mental imagery*. As such, the connection to imagination (and hallucination) is evident. Propo-

sitional memory, by contrast, is a mental state with semantic content. Nothing here indicates that proposition-imagination is involved: propositional memory necessarily involves neither the propositional attitude of acceptance nor the concept of possibility. I am unaware of any arguments in the literature in favor of the idea that proposition-imagination *is* necessarily involved in propositional imagination. Even a dissenting voice like Bernecker (2009, Ch.3), who argued against the widespread idea that propositional memory necessarily involves knowledge or belief, does not argue that it *does* necessarily involve imagination. (Of course, one *can* remember that p and imagine that p, i.e. occurrently accept that p is possible, at the same time.)

In conclusion, there are two types of memory: episodic memory and propositional memory. These two types are remarkably distinct with respect to their relation to imagination: episodic memory (2.37) is a type of action-imagination constrained by the past, and propositional memory (2.41) is a type of occurrent *believing* (or knowing) constrained by the past, which is logically independent of imagination.

2.7 Recapitulation

This marks the end of my analytic project of explicating imagination and allied concepts. In Section 2.2, Figure 2.1, I presented a schematic overview of the explicated concepts and their inter-relations discussed in this paper. I next briefly recapitulate the results of my inquiry.

I explicated vision, or visual perception, (2.1) and optical illusions (2.2) in the Conceptual Basis. I then began by pointing out a Divide among philosophers analysing imagination, between Imagers (2.4), who require mental imagery for imagination, and Wideheads (2.5), who permit imagery but do not require it. I then explicated proposition-imagination (ImProp) as occurrent acceptance of the possibility of proposition p, for some appropriate modality type (2.14). The connection between imagination and possibility is underwritten by all; my contribution is adding

the active and encompassing propositional attitude of *acceptance*, and putting the modality type as a 'parameter', whose 'value' is determined by p and the context in which it is imagined. I discerned, besides ImProp, also entity-imagination (ImEnt) and action-imagination (ImAct). I subdivided ImEnt in a sensory type (2.6) and a conceptual type (2.8); the sensory type I submitted as an instance of ImAct, and the conceptual type as an instance of ImProp. I then explicated three propositional attitudes closely connected to ImProp, supposition (2.15), counterfactual thought (2.16) and conceiving (2.21), as distinct from ImProp only in terms of the epistemic purpose with which we imagine the proposition; and, for the case of counterfactual thought, as implying a disbelief. I then explicated visualisation, for propositions (2.25) and entities (2.26), as mental states of imagination with representing visual content, and I explicated picturing, for proposition (2.28) and entities (2.29), as accurate visualisation restricted to concrete entities.

ImAct (2.31) I subdivided in two types recognised by all, inside-action-imagination and outside-action-imagination (2.30). I then provided explications of action-visualisation (2.32) and action-picturing (2.33), and I noted that visually imagining an entity is equivalent to imagining seeing that entity and to picturing that entity. Finally I explicated mnemonic ImAct (2.35) and related it to two main types of memory: episodic memory (2.37) and propositional memory (2.41). Rather surprisingly, it turns out that only the former involves imagination.

To bring it all together, I obtain the explication for a mental state of imagination *per se* by taking the inclusive disjunction of proposition-imagination (2.14) and action-imagination (2.31):

Imagination: Subject S has *mental state of imagination* m iff S imagines that p (2.14) or S imagines ϕ-ing (2.31), or both.	(2.42)

In all my explications I have used pretty standard concepts from various branches of philosophy, and this in combination with the content of my

explications has yielded a tight logical reticulum of imagination-concepts the world has never seen before — or so I imagine.

2.8 Imagination in practice — and in science

To conclude this Chapter, I next discuss some practical peculiarities to imagination that are important for understanding imagination but which are under-illuminated by the conceptual analysis of imagination that I have undertaken in this Chapter until now. I first clarify Yablo's notion of "accompaniment" of one type of imagination with another, as discussed in Section 2.3.2. I then make some comments on the cognitive-scientific perspective on imagination, which will be important for doing *epistemology* of imagination, which is the topic of the next Chapter of this Thesis.

2.8.1 On "accompaniment"

To begin, I wish to clarify Yablo's notion of "accompaniment" of one type of imagination with another. Recall the passage from Yablo (1993) that I quoted in Section 2.3.2, page 32:

> Imagining can be either *propositional* [ImProp] — imagining that there is a tiger behind the curtain — or *objectual* [ImEnt] — imagining the tiger itself. [...] To be sure, in imagining the tiger, I imagine it as endowed with certain properties, such as sitting behind the curtain or preparing to leap; and I may also imagine *that* it has those properties. So objectual imagining has in some cases a propositional accompaniment.

I have reduced entity-imagination (Yablo's "objectual imagination") to a propositional sub-type, conceptual entity-imagination (2.8), and a sensory sub-type, sensory entity-imagination (2.34), but Yablo's remark on "accompaniment" remains relevant: when you imagine an entity, and when you next find yourself imagining *that* this entity is such-and-so, then your

imagined entity has, in Yablo's terms, a "propositional accompaniment". We already encountered this case in the previous Section when I explicated *action*-imagination (of which sensory entity-imagination is a sub-type): when S imagines an action, and when the propositional attitude of acceptance that is dispositional in action-imagination (2.31) becomes *occurrent*, then S also imagines a proposition — this proposition "accompanies" the imagined action. So, this sense of Yablo's "accompaniment" has been well-accounted for by my explications of ImProp (2.13) and ImAct (2.31).

But there is more. In Section 2.4.2, feature VIII. *Under-determination*, I discussed how the content of proposition p *under-determines* the content of the mental state of imagining that p: when you imagine that p, it suffices to imagine *any* possible world that makes p true, of which there are generally infinitely many, and the choice of *which* world one imagines is under-determined by p. Yablo (1993, p.13, my emphasis) makes this same point, immediate following the quote above:

> Objectual imagining, I said, may be accompanied by propositional imagining. But it is the other direction that interests me more: propositional imagining *as accompanied by, and proceeding by way of,* objectual imagining. To imagine that there is a tiger behind the curtain, for instance, I imagine a tiger, and I imagine it as behind the curtain. Quite possibly though I imagine the tiger as possessed of various additional properties — facing in roughly a certain direction, having roughly a certain color, and so on — and I imagine besides the tiger various other objects — the curtain, the window, the floor between them — all arranged so as to verify my imagined proposition. In short I imagine a more or less determinate *situation* which I take to be one in which my proposition holds.

When we intend to imagine a proposition, Yablo notes, we often do so by imagining a *situation* — an entity — that we take to[54] *make true* the

[54] Yablo's phrase "takes to" is ambiguous and can be translated into *belief* or *acceptance*. Earlier, in Section 2.5.3, I noted that Yablo presumably understands it as belief. I suggest understanding it as acceptance, which seems much less problematic than belief, but can still do the required work here.

imagined proposition. This rings true. So, even though I — like all others — make a *conceptual* distinction between several types of imagination, it should already be evident even from these armchair considerations that these types will often come hand-in-hand in practice. But things are not as straightforward as they may seem. There are intriguing nuances in the notion of "accompaniment" that must be acknowledged. I mention three.

Firstly, I note that the notion of "accompaniment" resides in a grey area of *(in)voluntariness* of mental imagery. In Section 2.4.2, feature II. *Voluntary, Deliberate*, I distinguished between *voluntary* mental imagery, which is required for 'imagistic' imagination (i.e. visualising, picturing and outside-action-imagination), and *involuntary* mental imagery, which I called hallucinations. The question now obtrudes: when I imagine a proposition p, and this mental state of proposition-imagination is *accompanied* by mental imagery which I take to represent a situation which makes p true, to what extent is this mental imagery *voluntary*, i.e. to what extent can we call this accompaniment *imagination* rather than *hallucinations*? I do not think that there is a principled answer to this question. The mental imagery that accompanies one's mental state of imagining that p surely admits of 'various degrees of voluntariness': some imagistic accompaniments may be accepted (hence truly imagined), some may be discarded, some may simply 'pop up' and disappear as quickly as they came, and some may linger or return despite conscious attempts at discarding them, in which case they arguably become more like Dostoevski's polar bear (a 'gentle hallucination') than genuinely *imagined* mental imagery. Decisions must be made: to remain congruent with the literature, I shall henceforth treat all mental imagery that *accompanies* a mental state of imagination as *imagined* mental imagery (i.e. as *voluntary*), notwithstanding the issue just described.

Secondly, it should be clear that both the *amount* and the *type* of mental imagery that accompanies an imagined proposition p depends not only on the individual subject (see below) but also depends strongly on the *referential content of the proposition* itself. I note two characteristics of p

that stand out in this regard, and which, presumably, strongly influence the amount and type of mental imagery that would accompany imagining that proposition: (i) whether p mentions concrete or abstract objects, and (ii) whether p denotes a static moment in time or a dynamic event. To illustrate, suppose that you are requested to imagine the following three propositions one by one:

p_1. There is a cat on the mat.

p_2. Quantum mechanics is locally causal.

p_3. The Tower of Pisa is falling over.

For me — and, I presume, for most of us — imagining p_1 or p_3 is accompanied by much *more* mental imagery than imagining p_2. This is likely because p_1 and p_3 refer to concrete objects while p_2 does not. But there are important differences to imagining p_1 or p_3 too: when I imagine p_3 I also imagine the *sound* of the Tower of Pisa falling over, and I even imagine chaos ensuing on the streets, whereas when I imagine p_1 I simply imagine a static image of a cat on a mat but nothing else. The difference here resides in the fact that p_1 denotes a static moment in time whereas p_3 denotes a dynamic event.

Thirdly, and perhaps most importantly, it should be evident that both the amount and the type of mental imagery that accompanies an imagined proposition p will vary significantly from subject to subject. Many subjective factors will influence this variance, but two factors stand out in particular: (i) the *experience* (i.e. *memories*) that an individual imaginer has with the content of p, i.e. with the *topic* of her imagination, and (ii) the *imaginative capacities* that that individual imaginer has. I comment on each factor in turn.

(i) Concerning the experience that an individual imaginer has with the topic of her imagination. Undeniably, both the type and the amount of mental imagery that accompanies imagining a proposition p depends on the experience that an individual imaginer has with the content of p. To

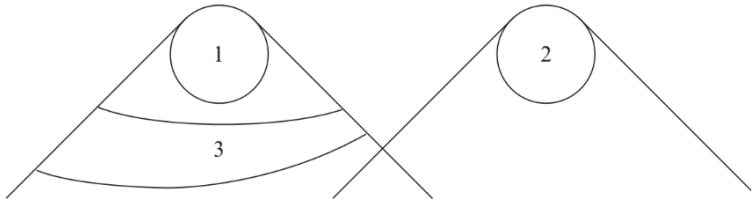

Figure 2.9: "Full specification of what happens in 3 makes events in 2 irrelevant for prediction about 1 in a locally causal theory." Image and caption from (Bell, 2001, pp. 224-225).

illustrate: if you have a cat at home, you may imagine *your* cat on the mat when imagining p_1, in all its splendid feline detail. This imagined cat will be very different from the cat that *I* imagine, which admittedly is a rather indeterminate cat. Concerning p_2, I find myself imagining John Bell's light-cone illustrations of locality when I imagine p_2 (see Figure 2.9), whereas you may imagine something entirely different.

It is important to realise at this point that mental imagery that accompanies imagining p *need not be mentioned, and, even, need not be implied* by the imagined proposition p, but may instead depend exclusively on the *experience* and the *associations* that an individual imaginer has with the content of p.[55] The diagram from Figure 2.9 that I imagine when I imagine that quantum mechanics is locally causal (p_2) is a case in point: this diagram is neither referred to, nor implied by, p_2; it is merely an artifact that I personally associate with p_2 due to the historical contingency that Bell's treatment of (non-)locality has been influential for my personal understanding of (non-)locality.

I note that the fact that mental imagery that accompanies proposition-imagination need not be mentioned by, and need not even be implied by, the imagined proposition p coheres well with my explication of proposition-

[55] Gendler (2008) calls such "chains of association" a subject's *aliefs* about p.

visualisation (2.25). Visualisations of propositions need not have anything in common with the topic of the visualised proposition; visualisations of propositions are not *pictures* of them. So, conceptually speaking, mental imagery that accompanies imagining a proposition can be regarded as *visualisations* of the imagined proposition — which, I submit, sounds very appropriate.

From these considerations, we can already conclude that imagined propositions can have highly non-trivial imagistic accompaniments, which I flag as an important observation for the *epistemology* of imagination; c.f. I return to it in Chapter 3. Getting ahead of myself, I even dare say that this observation suggests that accompaniments can be unplanned and unexpected and may thus even be genuinely *surprising* to the individual imaginers themselves, and may teach the imaginer something new about their *own* background beliefs and memories, etc. which certainly has important epistemological consequences; again, see Chapter 3.

(ii) Concerning the imaginative capacities of individual imaginers. It is an extremely interesting and well-documented fact about the human imagination that our *imaginative capacities* vary significantly from individual to individual. Some of us have an extremely vivid imagination (*hyperphantasia*), some tend to mingle and mix various sensory modalities in perception and imagination (*synesthesia*), and some even report a total *inability* to have or experience any mental imagery at all (*aphantasia*). This obviously has direct consequences for the "accompaniment" of one type of imagination with another in practice.

To understand these consequences, we must turn to cognitive science. So far, my analytic project proceeded predominantly by conceptual analysis. This is of course only a limited perspective on imagination. In the past decades, imagination has been extensively studied empirically. Several well-established regularities bear directly on the notion of "accompaniment", on the fact that individual imaginers have wildly varying imaginative capacities, and even on the connection between imagination and observable behavior. In the next Section, I discuss three of these

regularities: (i) the correspondence of eye-movements in perception and 'imagistic' imagination; (ii) the intuition that inside-action-imagination typically *triggers* outside-action-imagination; (iii) and the phenomenon of *aphantasia* and the epistemological questions it raises about imagination.

2.8.2 Notes on the cognitive science of imagination

(i) On eye-movement during 'imagistic' imagination

In this Chapter, I have treated imagination as a purely 'mental' concept. Imagination is the 'mind's eye', as they say. Notably, in the Conceptual Basis (Section 2.2), I began by noting that imagination and its allied concepts that I have explicated in this Chapter are not overtly and consistently connected to distinguishing observable behavior that can be judged — and *studied* — from a 3rd-person perspective. While this view is correct to a large extent (large enough for me to proceed with my 'armchair inquiry') it is also true that imagination, and notably also memory, are not *always* purely mental affairs. Often the body *influences* and *assists* the mind in imagining or remembering.

Concerning memory, it is known, for example, that *good body posture* facilitates the retrieval of episodic memories (Dijkstra et al., 2007); that *bad* body posture increases "depressive memory bias" (Michalak et al., 2014; Peper et al., 2017); and that "simple motor actions affect how efficiently people retrieve emotional memories", as e.g. *upward* body movement helps retrieve *positive* emotional memories faster (Casasanto and Dijkstra, 2010).

So too does the body often participate in imagining. Steier and Kersting (2019); Kersting et al. (2021) studied how young students extensively use hand and body-movements to *assist* their imagination (and to assist communication about their imagination to fellow students) when learning and attempting to comprehend complicated concepts from modern physics. Indeed, from the perspective of *embodied cognition* (Shapiro and Spaulding, 2021), it has been argued that body posture and movement

constrains[56] and *influences* the content of our imagination, and, consequently, influences what and how we *learn* through imagination (Rucińska and Gallagher, 2021). From this perspective, it is perhaps no surprise that imagination is explicitly appealed to in *dance therapy*, through what is known in Jungian psychology as "active imagination" (Chodorow, 1991; Chodorow and Jung, 2015; Wilde, 2011; Davis, 2019).

But perhaps the most direct connection between imagination and observable behavior is the well-established regularity that types of 'imagistic' imagination such as visualisation (2.25), picturing (2.28) and outside-action-imagination (2.30) *are typically*[57] *accompanied by eye-movements*, in the sense that our eyes move while we imagine 'imagistically' in roughly the same way as they would if we were to *perceive* the imagined scenario; see e.g. (Mast and Kosslyn, 2002; Rodionov et al., 2004; Sprenger et al., 2010; Laeng et al., 2014; Pathak et al., 2023), see also (Nanay, 2016b) and references therein, but c.f. (Pounder et al., 2022, p.188) and references therein. When we visualise a scenario, our attention — and, consequently, our gaze — turns to the *salient* features of this scenario: if the salient feature of the imagined scenario is, say, in the top-left of the scenario, then our eyes often also look top-left — just like our eyes would do if we were to actually *perceive* the scenario and pay attention to its salient features in the top-left.

While the rough correspondence between 'imagistic' imagination and eye-movements is well-established, it is not clear what the consequences of this correspondence are or should be for our understanding of imagination. Nanay (2016b) uses this correspondence to argue in favor of a particular account of the *content* of mental imagery. I mentioned this argument in footnote 33, p.34, but I also noted that the details of this argument are irrelevant for the purpose of this Chapter and well beyond its scope. Moreover, Laeng et al. (2014, §5) discuss with remarkable nuance how this correspondence seems compatible with various conflicting accounts of

[56] See next Chapter for more on constraints on imagination. [57] I.e. "probabilistically", see (Laeng et al., 2014) for a nuanced, detailed account. See also below.

mental imagery, and even with conflicting account of *cognition*.

But there is one plausible interpretation of this correspondence that I find particularly interesting and which is directly relevant for the other results from the cognitive-scientific perspective on imagination that I shall discuss next. According this interpretation, the correspondence between eye-movements and 'imagistic' imagination indicates that mental imagery causes our brain to *anticipate* further perceptions. On this, Laeng et al. (2014, p.278) write:

> According to Neisser (1976, pp.130–131), "the experience of having an image is just the inner aspect of a readiness to perceive the imagined object" and so that imagining and seeing are "only parts of a perceptual cycle" and under the control of "plans for obtaining information from potential environments." Within this account, imagery could be an anticipatory phase of perception like a "disposition to see" (see also Freyd (1987); Grush (2004); Kosslyn and Sussman (1995); Ryle (1949)) that takes place all the time; only when the perceptual pickup of information is either interrupted or delayed, [mental] imagery becomes subjectively experienced. Thus, in Neisser's account, the role of eye fixations during imagery seems particularly relevant, since anticipating visual information can guide gaze to the likely locations where this information will be found; c.f. Vickers (2007).

According to this interpretation, mental imagery causes the brain to *anticipate further perceptions* just like 'ordinary' perceptions do. The fact that 'imagistic' imagination is accompanied by 'perception-like' eye-movements is thus explained as directing our eyes into the direction where attention will most likely be *needed*, if the mental imagery were indeed followed-up by actual perceptions. This interpretation coheres well with the similar (but less often discussed) regularity that our pupils typically enlarge and dilate in ways that correspond to the *size*, *distance* and *brightness* of the imagined scenario (Laeng and Sulutvedt, 2014; Sulutvedt et al., 2018).

Perhaps one now wishes to include 'perception-like eye-movement' (preferably in more exact form) as an additional condition in explications

of 'imagistic' imagination-concepts such as visualisation (2.25), picturing (2.28) and outside-action-imagination (2.30). While I admit that this is tempting, as it would connect imagination to *observable* behavior, I am inclined to resist, for two reasons.

Firstly, it seems that eye-movement corresponds to 'imagistic' imagination only *typically*, but not necessarily. I submit that more empirical results are needed before we can make any necessity-claims about this correspondence. Secondly and relatedly, it appears that the *strength* of the correspondence varies significantly between individuals and notably depends on the *memories* that the imaginer has of the imagined scenario. Laeng et al. (2014, p.263) write, for example:

> [W]e predicted that when observers looked at an empty screen and at the same time generated a detailed visual image of what they had previously seen, their gaze would probabilistically dwell within regions corresponding to the original positions of salient features or parts. Correlation analyses showed positive relations between gaze's dwell time within locations visited during perception and those in which gaze dwelled during the imagery generation task. Moreover, the more faithful an observer's gaze enactment, the more accurate was the observer's memory, in a separate test, of the dimension or size in which the forms had been perceived.

While I find this regularity intriguing, I note that it prevents us from including eye-movement in our explications of 'imagistic' imagination, because we can visualise things we have little to no memories of. In cases where our memories play a *negligible* role in visualisation, the 'perception-like' eye-movement no longer serves as a useful *observable* distinguishing condition between different 'imagistic' imagination concepts, and it certainly is no *necessary* condition for having a mental state of 'imagistic' imagination. I repeat that further empirical results are required before we can incorporate a condition pertaining to eye-movement in our explications of 'imagistic' imagination-concepts; c.f. Pounder et al. (2022).

(ii) On "triggering" outside-action-imagination

In Section 2.6, I distinguished two types of action-imagination (2.30): *outside*-action-imagination and *inside*-action-imagination. Following Jeannerod (1994), I distinguished between these two types on the basis of their respective content: *outside*-action-imagination necessarily involves *sensory* content, and *inside*-action-imagination necessarily involves *motor* content. This distinction is well-grounded in the literature; see e.g. (Lacey and Lawson, 2013; Kilteni et al., 2018; Pearson, 2019; Nanay, 2021)

But, whereas outside- and inside-action-imagination are conceptually distinct and each type is grounded in different parts of the brain just like ordinary motor and sensory content are grounded in different parts of the brain (Pearson, 2019; Kilteni et al., 2018), it seems that these two types of action-imagination often *accompany* each other in practice. Try imagining *reaching* for an apple (inside-action-imagination) without *visually* imagining yourself reaching for an apple (outside-action-imagination). Perhaps you can do it if you try hard enough (or if you are an *hyperphantasiast*, see below), but I cannot.

Recently, Kilteni et al. (2018) established a result that corroborates the intuition that the two types of action-imagination are strongly connected in practice: having a mental state with endogenous motor imagery typically *involves predicting the sensory consequences of the imagined movement*. In other words, mental states of inside-action-imagination, i.e. mental states with endogenous motor content, typically *activate* and *prepare* our sensory processing mechanisms for what should come if we were to actually execute the imagined action. Kilteni et al. (2018) thus demonstrated that our brain deals with *endogenous* motor content in much the same way as it deals with *exogenous* motor content: both *activate* and *prepare* our sensory processing mechanisms in much the same way. (Note how well this result harmonises with the correlation between mental imagery and eye-movement that I discussed in the previous Section.)

At first sight, Kilteni et al.'s result supports — nearly *explains* —

the intuition that inside-action-imagination is typically accompanied by outside-action-imagination in practice. Surely, if one *expects* some incoming sensory input, then one is very close to *imagining* this sensory input too. But we must be careful not to draw this conclusion too quickly. If one expects some incoming sensory input, one is also close to *(gently) hallucinating* the sensory input. Recall my discussion of Dostoevsky's polar bear (Section 2.4.2, p.47): once we have the idea of a polar bear in mind, we often find ourselves generating mental imagery of polar bears involuntarily, even upon the explicit request not to do so. The difference between imagination and hallucination resided here whether or not the mental imagery is *voluntary*. The results of Kilteni et al. (2018) show us yet again that the notion of voluntariness must be handled with care. I submit that further empirical results pertaining to the voluntariness of this mental imagery are required before we can draw conclusions about the practical connection between motor imagery and *imagination*.

So much for "accompaniment" of one type of imagination with another. To conclude this Chapter, I turn to a phenomenon that has received much attention in science and philosophy in the past few years: *aphantasia*.

(iii) On the phenomenon of *aphantasia*

Mental imagery plays deeply important roles in most of our lives — in our memories and in our imagination, and even in our dreams. But anyone who attempts to *explain* where, how, and why mental imagery plays such important roles in our lives, has to deal with the fact that individual imaginers have wildly varying imaginative capacities. As I wrote at the end of Section 2.8.1, some of us have an extremely vivid imagination (*hyperphantasia*), some tend to mingle and mix various sensory modalities in perception and imagination (*synesthesia*), and some even report a total *inability* to have or experience any mental imagery at all (*aphantasia*). In this Section, I make some brief but necessary comments about this latter phenomenon of *aphantasia*.

Some individuals lack the ability to experience mental imagery. Al-

though this phenomenon was already mentioned in one of the earliest empirical works on mental imagery (Galton, 1880), our understanding of this phenomenon has long been based predominantly on self-report; c.f. Pearson (2019). This changed recently, when the phenomenon got a name — *aphantasia* — and became the topic of empirical research; e.g. Zeman et al. (2015, 2016); Keogh and Pearson (2018); Dawes et al. (2020); Bainbridge et al. (2021); Pounder et al. (2022); Dupont et al. (2022). I next highlight two particularly remarkable results from this body of research.

For a long time, it was an open question whether aphantasiast are unable to *have* mental imagery, or whether they are able to have mental imagery (like everyone else) but they are somehow unable to be *aware* of their own mental imagery — i.e. that aphantasists have "very poor metacognition" (Keogh and Pearson, 2018, p.58). Intuitively, the latter explanation is more plausible than the former. But, strikingly, current results suggest that the former explanation is true: it appears that aphantasiasts are genuinely *unable to have* mental imagery. Keogh and Pearson (2018) demonstrated, for example, that aphantasiasts (i.e. subjects who self-report being unable to experience mental imagery) actually perform below average on tasks that crucially employ mental imagery. Dawes et al. (2020) showed that self-proclaimed aphantasiasts report "less vivid and phenomenologically rich autobiographical memories and future imagined scenarios",[58] and even that they report "fewer and qualitatively impoverished dreams compared to controls" (p.1). It thus apppears that aphantasia bears not only on *imagination*: aphantasia bears on *all types of* endogenous mental content, including the content of memories and dreams (i.e. hallucinations); recall my conceptual geography presented in Figure 2.1.

Even though there is reasonable consensus in the literature that aphantasiasts are genuinely unable to *have* mental imagery, it should be clear that it is not the case that aphantasiasts are unable to imagine *at all*.

[58] This further supports the 'continuity hypothesis' that imagination and memory are inter-related rather than fundamentally distinct; recall Section 2.6.3.

In terms of the concepts explicated in this Chapter, the above-mentioned results suggest that aphantasiasts are only unable to e.g. visualise (2.25) or picture (2.28) propositions, and that they are unable to imagine *actions* 'from the outside' (2.30). As all above-mentioned authors note, however, the empirical research on aphantasia is still recent and limited in both size and scope, and many more further results are needed before we can draw far-reaching conclusions. Notably, I submit, it would be very interesting to see how aphantasiasts exhibit the two phenomena (i) and (ii) discussed above. To repeat: (i) on the correlation between eye-movement and mental imagery — would this be absent for aphantasiasts? See e.g. Pounder et al. (2022), who note that current results are too unclear to draw conclusions. (ii) On the idea that inside-action-imagination triggers outside-action-imagination — would this not occur for aphantasiasts, and can aphantasiasts even have *motor* imagery? The research from Dupont et al. (2022) suggests that aphantasiasts *cannot* conjure up endogenous motor content, thus suggesting that aphantasiast are unable to imagine action from the outside *and* from the inside. Again, however, further results are needed before we can draw conclusions.

But there is one surprising empirical result pertaining to aphantasia about which there already is reasonable consensus in the scientific literature: while aphantasiasts perform significantly below average on tasks that crucially employ mental imagery, they tend to perform *above average* on *spatial reasoning* tasks; see e.g. (Keogh and Pearson, 2018; Pearson, 2019); but c.f. Pounder et al. (2022) for reservations on this.

This result is surprising because, intuitively, in performing spatial reasoning-tasks, we often crucially employ mental imagery. Indeed, in the next Chapter, I shall argue that one of the reasons why imagination deserves to be called a 'distinctive source of knowledge' is precisely because we can perform spatial reasoning tasks *in the imagination* — that is, by forming a mental image of some scenario and reasoning spatially *about this imaginary scenario*. Aphantasiasts would be unable to do this. Thus, the fact that aphantasiasts can perform above average on spatial

reasoning tasks suggests that spatial reasoning need not *necessarily* be based on mental imagery, but that it can also be performed 'more abstractly', without employing mental imagery. Indeed, these results seem to suggest that visualisation is only *one of many ways* in which we can perform an epistemic task, even epistemic tasks that appear to crucially employ mental imagery (Keogh et al., 2021).

Given that, for *most* of us, mental imagery plays a deeply important *epistemic* roles in our lives, it remains an intriguing question how it is that aphantasiasts can have *normal* epistemic lives (Keogh et al., 2021; Arcangeli, 2023). This question is answered partly by the above-mentioned fact that aphantasiasts *can* perform well on spatial reasoning tasks. It is also answered partly by the fact that aphantasiasts are not unable to imagine *at all*: they are only unable to conjure up mental imagery. Aphantasiasts can imagine propositions (2.12) perfectly well — as such, they can conceive (2.21), suppose (2.15), and reason with counterfactuals (2.16). Indeed, as Keogh et al. (2021, p.277) write:

> Aphantasic individuals can also be highly imaginative and are able to complete many tasks that were previously thought to rely on visual imagery, demonstrating that visualization is only one of many ways of representing things in their absence. The study of extreme [variations in] imagination reminds us how easily invisible differences can escape detection.

I repeat once more: further empirical results are required before we can draw conclusions about imagination and its sub-types on the basis of cognitive-scientific research on imagination. Until then, I submit, philosophical 'armchair inquiry' such as conceptual analysis of imagination remains a valid and useful way of analyzing imagination, both for increasing our understanding of the *concept* of imagination and its sub-types — as I did in this Chapter — and for increasing our understanding of the *epistemic value* of imagination and its sub-types — to which I shall turn in the next Chapter.

2.9 Conclusion

To conclude and recapitulate once more, in this Chapter, I distinguished and explicated two types of imagination (2.42): proposition-imagination (2.14) and action-imagination (2.31). I also explicated the closely-related concepts of perception (2.1), optical illusion (2.2), supposition (2.15), counterfactual thought (2.16), conceiving (2.21), proposition-visualisation (2.25) and action-visualisation (2.32), and proposition-picturing (2.28) and action-picturing (2.33). Additionally, I explicated mnemonic imagination (2.35), i.e. a mental state of imagination with mnemonic content. I then argued that one of the two main types of memory, episodic memory (2.37), *is* mnemonic imagination; and I noted that, somewhat surprisingly, the other main type of memory, propositional memory (2.41), does *not* seem to be a type of imagination. See Figure 2.1 for an overview of the logical connections between all these explicated concepts.

I then discussed Yablo's notion of "accompaniment" (Section 2.8.1), which denotes the cases where having a mental state of one type of imagination is, rather automatically, followed by mental states of other types of imagination. Finally, I commented on the cognitive-scientific perspective on imagination (Section 2.8.2), and concluded that this perspective does not conflict with the explications of, and logical relations between, the concept of imagination and related concepts, that I proposed in this Chapter.

So much for my Carnapian project of explication. I next turn to the question how imagination functions as a *source of knowledge*.

Chapter 3

Knowledge Through Imagination

3.1 Introduction

When it comes to the imagination, not even the sky is the limit. Above and beyond the sky, the imagination runs wild, without limits and with exuberance. *We can imagine anything we want.* The limitlessness of imagination is, epistemically speaking, both a virtue and a vice. Whereas imagination is an amazing and widely celebrated source of new ideas, hypotheses, models and theories (viz. the 'context of discovery' of philosophy of science), these can never be *justified* by imagination alone but must, in order to attain the laudable status of *knowledge* of the world, be justified firmly by reason and observation (viz. the 'context of justification'). Whence the widespread claim that imagination cannot be a source of knowledge: imagination seems impotent when it comes to epistemic justification. To imagine that p and to know that p seem contraries.

Notwithstanding, a central question in the epistemology of imagination that is debated with increasing fervour is:

> **The Question of Knowledge Through Imagination:**
> Is imagination a source of knowledge of the natural world?

Answers to this Question range from "absolutely, self-evidently not" to "of course, we gain knowledge through imagination all the time", with many positions taken up in between these two extremes. The majority of the relevant literature was published only in the past few decades, so the debate is still rapidly developing. In this Chapter, I aim to contribute clarity and structure to this debate about the Question of Knowledge Through Imagination as follows.

In Section 3.2, I begin with some terminological preliminaries, and I specify that my explanatory target in this Chapter is imagination as a source of *quasi-perceptual knowledge*. In Section 3.3, I elaborate on the concept of quasi-perception and explicate it (3.7). I then introduce *two-step schemas* for describing perceptual (3.8) and quasi-perceptual (3.9) belief-yielding processes. I then discuss three examples, and I argue at length why quasi-perceptual belief-yielding processes require (what I call) *meta-beliefs* about the accuracy of our imaginings.

In Section 3.4, I discuss the justification of quasi-perceptual beliefs. I begin with providing an explicit criterion for when quasi-perceptual beliefs are justified (3.10). I then discuss the *Constraint Claim* (3.11), which is the claim that imagination can be a source of justified beliefs when the content of our imagination is *properly constrained*. I then respond directly to Kinberg and Levy (2022), who put forward an interesting but, in my view, unconvincing dilemma for proponents of the Constraint Claim.

Finally, in Section 3.5, I discuss in which sense it can be said that imagination is a 'source of knowledge'. I first show that that imagination is not a so-called *basic* source of knowledge (3.15): imagination can not yield knowledge 'on its own'. I then argue that imagination is (what I call) a *crucial* source of knowledge (3.16): there exist knowledge-yielding processes where imagination is at least partially responsible for *both* the formation of the belief *and* its justification. On the basis of this discussion, I then favorably review arguments in favor of the idea that imagination is a source of *otherwise-inaccessible* knowledge (3.17): imagination can yield knowledge that other sources of knowledge do not have access to.

3.2 Preliminaries

3.2.1 Acts of imagination

In the previous Chapter, I proposed explications for the mental state of imagination (2.42) and its various types and sub-types. The two explicated types of imagination were proposition-imagination (2.14) and action-imagination (2.31), which then divided into many more sub-types (inside and outside action-imagination, supposition, visualization, picturing, etc.; summarised in Section 2.7, Figure 2.1). This clarified the relation between imagination and other mental states such as perception, belief and memory. The Question for this Chapter is what role mental states of imagination may play for us in gaining knowledge of the world.

To tackle this Question, it does not suffice to look at imagination only as a mental state. Knowledge is gained through imagination not in a single instance, by means of a single mental state of imagination; it is gained through a mental episode, which is a *project* that takes some effort and time. Multiple different mental states of imagination (and other types of mental states) will be involved in these episodes, and these mental states can relate to each other in interesting, non-trivial ways. Hence we must focus our attention not on single mental states of imagination (and otherwise) but instead at the mental *process* that these mental states figure in: e.g. at what happens in the *performance* of a thought experiment.[59] Stuart (2021, p.1332) concurs:

> If we are to do state-based epistemology of imagination, that is, if we are to find out how imagined mental states come to be known or play a role in gaining new knowledge, I suggest that the imagined content must figure somehow into an argument, inference, or other kind of *process*.

I shall call such an imagination-based mental process an *act of imagina-*

[59] Imagination is also occasionally connected to physical action, as in the cases of pretence, acting, stage-performance or playing games of make-believe. Such behavior is beyond the scope of this Chapter: it will be discussed extensively in the next Chapter.

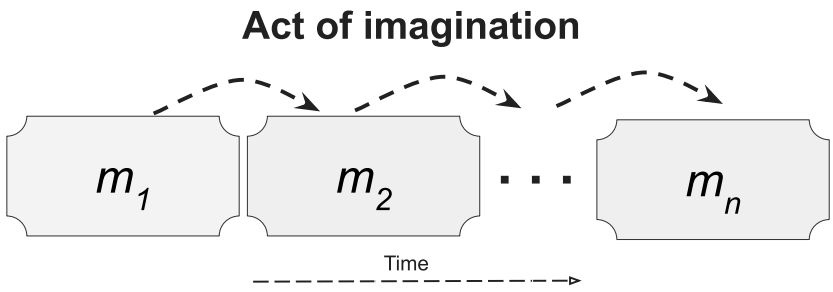

Figure 3.1: Visualisation of an act of imagination.

tion. All acts of imagination that I consider in this Chapter are acts of imagination performed with an epistemic purpose: they are *epistemic acts of imagination.* For the sake of convenience, I shall call them simply acts of imagination.[60]

To set things up in a manageable way, I shall follow Langland-Hassan (2016) and simply think of acts of imagination as *sequences* of mental states of imagination: $A := (m_1, m_2, \ldots, m_n)$. See Figure 3.1. I shall assume that acts of imagination are temporally ordered and that the sequential mental states of imagination *hang together* content-wise, in the sense that the sequential mental states are, at least to a significant extent, *coherent and inter-related*, notably to the extent that they share all or some of their intentional objects, which all exist in the same possible worlds of the same type of modality. Explication:

> [Act of Imagination] Subject S performs act of imagination $A := (m_1, m_2, \ldots, m_n)$ iff S has a sequence of mental states of imagination $(m_1, m_2, \ldots, m_n)$ that are temporally ordered and which share all or some of their intentional objects, which all exist in the same possible worlds of the same modality type. (3.1)

The type of imagination relevant for this Chapter is *action-imagination*

[60] In the literature, acts of imagination are also known as "imaginative episodes" (Langland-Hassan, 2016; Kinberg and Levy, 2022), but I prefer the term 'act of imagination' over 'imaginative episode' because the former emphasises the *deliberacy* of the imaginative process — it is "purposive" (Dorsch, 2015) *mental action* — which is important, epistemologically speaking.

(2.31). Recall that I explicated action-imagination as the inclusive disjunction of *inside*-action-imagination and *outside*-action-imagination (2.30):

> [ImActOut] Subject S *inside-imagines* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate type of modality, and S has an occurrent endogeneous mental state such that its sensory content represents the event of S ϕ-ing.
>
> [ImActIn] Subject S *outside-imagines* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate type of modality, and S has an occurrent endogeneous mental state such that its motor content represents the event of S ϕ-ing.

(3.2)

All acts of imagination (3.1) discussed in this Chapter involve (at least) sequences of mental states of action-imagination (3.2).

I next make explicit an important simplification that too often remains implicit in debates about the Question of Knowledge Through Imagination: I assume that, while performing an act of imagination, the imagining subject does not receive new perceptual input that is relevant for the epistemic purpose at hand. In other words, I shall consider only acts of imagination that are performed *in absence of relevant perceptions*. This importantly excludes from consideration acts of imagination where the imaginer uses the objects in its direct environment to *assist* its act of imagination; think of acts of imagination with immediate practical relevance in some real-world scenario, e.g. when you look at a cliff and imagine climbing it, and then climb it as imagined, c.f. Williamson (2016). Such acts of imagination that involve relevant perceptions will be discussed thoroughly in the next Chapter. The acts of imagination that I discuss in this Chapter can all be performed by sitting down, closing one's eyes and ears, and just *imagining* something, without external input: the sensory mental states are purely *endogenous*.

Having made this simplification, it is important to re-emphasise that, while one performs an act of imagination, one's mental state need not be

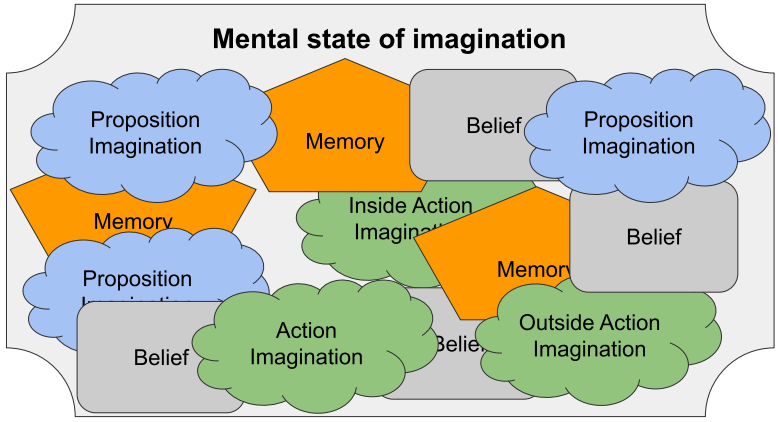

Figure 3.2: Visualisation of a complex, conglomerate mental state of imagination-memory-and-belief.

Figure 3.3: A complex jungle. (Image courtesy of istockphoto.com.)

a mental state of *only* imagination. The mental states considered may have other types of content and attitudes as well, notably the content of *memories* and *beliefs*. We can *and do* remember and believe things while we perform acts of imagination; and vice versa. As such, the mental states

that make up an act of imagination — and typically are — *conglomerate* mental states. Here lies a core task for the epistemologist of imagination: to reconstruct and dissect acts of imagination that consist of complex, conglomerate mental states, and to identify and disentangle the contribution of *imagination* to the acquisition of knowledge, *vis-à-vis* the contribution of e.g. memory and beliefs. See Figure 3.2 for an illustration of a complex, conglomerate mental state. If this Figure seems cluttered, that is because it *is* cluttered. But this does not prevent us from identifying and disentangling its distinct components, just like we can identify and disentangle the distinct components of a complex jungle; compare Figure 3.2 to Figure 3.3. Recognizing this will be crucial for understanding how imagination can be a source of knowledge.

3.2.2 Knowledge

I shall limit my attention to *propositional knowledge* of the world, a.k.a. knowledge-that-p, where p is a proposition about contingent or necessary features of the natural world; henceforth simply referred to as "knowledge". I proceed with the standard explication of knowledge as *justified true belief*:

> [Knowledge] Subject S knows that p iff S believes that p, p is true, and S's belief that p is justified. (3.3)

I shall take the concepts of truth and belief for granted; c.f. Chapter 2, Section 2.2. Concerning the *epistemic justification* of beliefs (henceforth simply referred to as "justification"), I follow the majority of the relevant literature and adopt the following reliability-and-robustness-criterion for the epistemic justification of beliefs:

> [Epistemic Justification] Subject S's belief that p is *justified* iff S obtained that belief through a *reliable* and *robust* process. (3.4)

Roughly speaking, a process is *reliable* iff it yields true beliefs more often than not; and a process is *robust* iff it would, more often than not, yield the same belief under *slightly* different circumstances. Much of this Chapter is directly concerned with the question whether *distinctively imaginative* processes (defined in Section 3.5.2) can be reliable and robust as such.

I note that I follow e.g. Dorsch (2016b); Kinberg and Levy (2022) (see below) and regard robustness, which guards knowledge against Gettier-style counterexamples, as a necessary condition for epistemic justification. I acknowledge that robustness is also regularly treated not as a necessary condition for justification but as an independent necessary condition for knowledge: knowledge as justified true *robust* belief. The difference is insubstantial for my present purpose: wherever robustness resides, the relevant question is only whether it is *there*.

The criterion for epistemic justification (3.4) that I adopt is not accepted by all. One may, for example, advocate a stricter criterion for epistemic justification that requires the additional condition that S is in a position to *know* that their belief was obtained through a process that reliably yields true and robust beliefs, i.e. S is in a position to know that condition (ii) in (3.4) is met. Fortunately, the distinction between the externalist and internalist criteria for justification is not crucial for my purpose: for the cases that I discuss in this Chapter, once the externalist-criterion for epistemic justification (3.4) is met, which by itself is hard enough, then the stricter internalist-criterion is often *also* met — this is the case because, as I shall argue, gaining knowledge through imagination requires having a meta-belief *about* the reliability of the process of knowledge-acquisition, which gets us rather close to meeting the above-mentioned additional condition for justification.

At this point, I wish to acknowledge that knowledge is not the only interesting epistemic product of imagination. Two notable other epistemic products of imagination are *understanding* and *conceptual change*. Both of these are highly interesting in their own right, and it has been argued in recent years that the predominant focus on (propositional) knowledge in

the epistemology of imagination is limiting our understanding of imagination in general; see notably the work of Stuart (2015, 2016, 2018, 2020) but see also e.g. Kuhn (1977); Nersessian (1999); Steier and Kersting (2019); Alstein et al. (2022); c.f. de Regt (2014, 2017, 2020).[61] I applaud this development, but I nonetheless focus my analysis on propositional knowledge in this Chapter — I shall directly respond to Kinberg and Levy (2022) at the end of this Chapter, whose target is propositional knowledge too.

I next distinguish four ways in which imagination is often discussed as a potential source of knowledge:

1. Imagination as source of *quasi-perceptual*[62] knowledge of the natural world: "we immerse ourselves in a scenario, trying to "live it" in our minds" (Kinberg and Levy, 2022, p.3) and we obtain beliefs "quasi-perceptually", meaning that "the presence of a mental image [plays] a crucial cognitive role in the formation of the belief" (Gendler, 2004, p.1152). The paradigmatic examples here are mental simulation and, more specifically, thought-experimenting (Stuart et al., 2018; Brown and Fehige, 2019).

2. Imagination as a source of *practical* knowledge, mainly in the form of (i) future-predictions with immediate relevance for action, and (ii) increased skill. Examples of (i) are looking at a dangerous cliff, imagining a specific way of climbing up without falling, and then climbing it on the basis of this act of imagination, or predicting whether you would enjoy living in the house that you're currently visiting (Williamson, 2016). An example of (ii) is the well-established practice of robustly increasing athletic performance through 'mere' mental simulation (Jeannerod, 1994; Dello Iacono et al., 2017). C.f. Lombrozo (2020); Aronowitz and Lombrozo (2020).

3. Imagination as a source of *modal* knowledge, notably knowledge of

[61] I am a co-author on Alstein et al. (2022). I will return to this article in Chapter 4. [62] This term is inspired by Sartre (1948); Gendler (2004), who call it *quasi-observational*. I take quasi-perception (3.7) to be the fundamental term. Per (2.1), then, quasi-observation is deliberate and attentive quasi-perception.

(im)possibilities and counterfactual relations (Brown, 1991; Ichikawa, 2016; Iranzo-Ribera, 2022; Berto, 2022). In this sense, imagination is closely tied to supposition (2.15), counterfactual thought (2.16) and conceiving (2.21); recall my explications of these concepts in the previous Chapter; c.f. (Arcangeli, 2019; Salis and Frigg, 2020). Perhaps unsurprisingly, it is sometimes argued that such "suppositional imagining does not raise novel epistemic questions, [...] epistemically speaking, it is plain old hypothetical reasoning" (Kinberg and Levy, 2022, p.3).

4. Imagination as *essential for* sources of knowledge: a *sine qua non* for the functioning of other, more 'traditional' sources of knowledge such as perception and reason. For example, ever since Kant and, more recently, Strawson (1970), it is argued that "most if not all perceptual experiences are infused with imagination" (Brown, 2018, p.133). In this sense, imagination is similar to memory, in that "what we think of as "our knowledge," in an overall sense, would collapse if memory [*and imagination*] did not sustain it" (Audi, 2005, p.74).

These four ways are inter-related. Notwithstanding, important differences pertain to the types of imagination involved and the types of knowledge gained in each — and to the *controversy* surrounding each. Epistemological analysis of imagination typically proceeds by focusing on only one of these four topics. This Chapter is no different: my topic is 1 — imagination as a source of quasi-perceptual (propositional) knowledge — although I must and shall occasionally refer to types 2–4.

3.2.3 Memory revisited

In Chapter 2, Section 2.6.3, I proposed an explication for *mnemonic imagination* (2.35), i.e. a mental state of imagination with mnemonic content, and I argued that the distinction between memory and imagination is not sharp but vague. I need not and shall not make further commitments

about memory and its relation to imagination — the explications of memory and of mnemonic content will remain the proverbial elephant in the room — but I do wish to add some further comments on the distinct *types* of memory that are discussed in the literature, specifically in relation to the question whether memory is a source of knowledge.

Just like there is no grand theory of imagination about which there is broad consensus available in the literature, so too is there no overarching theory of memory (Michaelian and Sutton, 2017). But — again, just like the case of imagination — there is reasonable consensus about there being different *types* of memory. In the previous Chapter, Section 2.6.3, I already discussed the distinction between *episodic* and *propositional* memories. I shall return to these two types below. But there is a deeper distinction within the concept of memory that I must turn to first.

I acknowledge the distinction between so-called *declarative* and *non-declarative* memory (Squire, 2009). Declarative memory is the conscious retrieval of facts and events that one experienced in the past: both episodic memory (2.37) and propositional memory (2.41) are types of declarative memory. Non-declarative memory, by contrast, is experience that shapes our skills, habits, dispositions, intuitions and personality traits — in short, *behavior* — "without requiring any conscious memory content or even the experience that memory is being used" (*ibid.*, p.12711). For example, the experience of being trampled by an elephant as a child yields the declarative memory of being trampled by an elephant, and it may yield the non-declarative memory of habitually fearing elephants. I shall limit my attention to declarative memories, i.e. to the conscious, deliberate and occurrent *use* of memory, just like I did in the previous Chapter. I make this same choice here partly because the reliability and robustness, i.e. justificatory force, of non-declarative memory as a source of knowledge is, understandably, highly dubious; c.f. the 'memory wars' from the 1990s (Crews, 1995). The epistemology of non-declarative memory is highly interesting, but it is sadly beyond the scope of this Thesis.

Within declarative memories, I again distinguish only between *episodic*

memories (2.37) and *propositional memories* (2.41). I repeat my explications of these two types of memory for the sake of convenience:

> [Episodic memory] Subject S *remembers* ϕ-ing iff S imagines ϕ-ing (2.30), S ϕ-ed in the past, and there is an appropriate causal connection from S's ϕ-ing in the past to the sensory and motory content of S's current mental state of imagination. (3.5)

> [Propositional memory] S remembers that p iff S knows that p, and S knew that p in the past, and there is an appropriate causal connection from S's past knowledge that p to S's occurrent knowledge that p. (3.6)

Just like the type of imagination most relevant for this Chapter is action-imagination (3.2), the type of *memory* most relevant for this Chapter is episodic memory (3.5). Just like I construe acts of imagination as sequences of mental states of action-imagination, I construe acts of episodic memory as sequences of mental states of episodic memory (3.5). For the sake of convenience, however, I henceforth refer to acts of episodic memory simply as episodic memory (the difference between the two is insubstantial for my present purpose).

Next, some general comments on memory as a source of knowledge. There are serious epistemological problems haunting memory — so serious, that these problems got their own entry in the Standford Encyclopedia of Philosophy (Senor, 2019), which notably is not the case for imagination. Many of these problems are informatively analogous to the epistemological problems surrounding imagination, and the relation between memory and imagination with respect to these epistemological problems is relatively under-explored in the literature. I briefly mention two problems for memory as a source of knowledge; I shall return to them when I discuss their imaginative counterparts in the next Sections.

Firstly, there is the problem of *novelty*, on which I shall spend the most time because it is directly relevant for the case of imagination. It

is occasionally argued that memory is not a source of knowledge because memory is not even a source of *novel beliefs*, let alone a source of knowledge. Memory, it is argued, *preserves* beliefs (and knowledge) that were obtained through some source other than memory, but it cannot *generate* them by itself. Consider, for example, Audi (2005, p.74-5), who writes:

> [S]urely one cannot know anything from memory without coming to know it [first] through some *other* source. If we remember it and thereby know it, we *knew* it, and we must have come to know it through, say, perception or reasoning.

Given that knowledge implies belief and justification, and noting that Audi (1995) famously argued that memory *is* a (generative) source of novel justification, we can conclude that Audi (2005) essentially argues in the above-quoted passage that memory is not a source of novel beliefs but that memory can only provide us with beliefs that we already *had*. As such, Audi (2005) calls memory a *preservative* source of beliefs (and knowledge), restricting its function to "retaining knowledge already gained" (*ibid.*, p.75), as opposed to a *generative* source.

This line of reasoning makes sense at first sight. If you saw an elephant in the room in the past, and the elephant is still there, then you can use memory to *retain*, but not *obtain*, the knowledge that there is an elephant in the room. To *obtain* this knowledge you would need to *perceive* the elephant in the room, for example. Notwithstanding, I wish to argue that memory *can be* a generative source of beliefs. To see how, we need to pay attention to the distinction between propositional and episodic memory: propositional memory is not a source of novel beliefs, but episodic memory *is*. Let me explain.

Like most authors, I understand propositional memory (3.6) to *be* the recollection of past knowledge — hence the recollection of past beliefs. If you did not know (hence believe) that p in the past, then you cannot remember that p in the present. This partly vindicates Audi's argument because *propositional* memory cannot generate novel beliefs.

But this is not the case for episodic memory (3.5). Episodic memory

is the recollection of past perceptions. And not nearly all our past perceptions have propositional form to the extent that they can be reasonably called *beliefs*; presumably, only a *tiny fraction* of the sum-total of our past perceptions can be said to have propositional form as such. The pool of content that episodic memory draws from is not a library of propositions, it is predominantly a messy cluster of 'experience'. Some of this 'experience' has propositional form, while much of the rest of it is just that: *experience*. Qualia, sensory and motory content, affective content, emotions and such. (This difference in mnemonic content is reflected precisely in the distinction between propositional and episodic memory.) Episodically remembering one's perception of a past event e implies that one *perceived* e in the past, but it does not imply that one *believes* everything that one can in principle believe about e.

So this is where episodic memory can serve as a source of novel beliefs: through episodic remembering, we can re-live past perceptions and associate with these perceptions novel propositions that can be *believed*. The processes of gaining novel beliefs through memory (and imagination) as such are often described as processes where we *make beliefs propositionally available*; see notably Gendler (2004, 2010); Lombrozo (2020); c.f. Mach (1960).[63] (I shall often use this phrase, but I note here that it is arguably more accurate to say that these are processes where we make propositions available *for* belief; see Section 3.3.) I next discuss two examples of this process of making beliefs propositionally available.

[63] One may insist that this mere "making propositionally available" of beliefs is not enough to consider memory as a source of novel beliefs. Perhaps one demands more novel *oomph*, the kind of novelty that only perception (and perhaps reason) can provide. If this is the case, then the discussion is over: memory is then just not a source of novel beliefs. The advantage of this position would be that perception (and perhaps reason) remains the only sovereign sources of truly novel beliefs (and knowledge), and epistemology can proceed as it always has. The disadvantage of taking up this position, however, is that we prematurely dismiss other potentially interesting sources of knowledge: not only memory, but also imagination. If memory and imagination cannot provide novel beliefs, then they cannot be sources of knowledge, and there is no phenomenon to be explained. But, I insist, like many others, that there *are* phenomena to be explained. I return to this issue in the next Section. See also (Gendler, 2004, p.1157, fn.7).

First, a rather trivial example. Suppose you saw an elephant in the room last week, and so you believe that there was an elephant in the room last week. Today you learned that African elephants have much larger ears than Asian elephants. You now decide to relive your perception of seeing the elephant in the room in your episodic memory and pay particular attention to the size of the elephant's ears. You find that the elephant in your episodic memory has remarkably small ears — certainly not remarkably large ones — and so you form the novel belief that you saw an *Asian* elephant in the room last week. This example is one where the subject obtains novel observational concepts in between the occurrence of the remembered event and the episodic memory itself, which may not be a convincing example of a genuinely *novel* belief (see e.g. (Miyazono and Tooming, 2023a, fn.3)), so I shall give another.

Second, a less trivial example. Suppose your friend plays a very simple rhythm for you on the drums: *bam bambam bambam bam*. He next asks you: how many times did I hit my drum? You did not count this while you were listening to the rhythm, so you have no answer readily available. You have no belief about this matter of fact. In order to come up with an answer, then, you replay the drumbeat in your memory and *count* the number of times the drum is hit. Thus you form the novel belief that your friend hit the drum six times. We would form a novel belief in a similar way to the question: how many times does Phil Collins hit the drums in his famous drum-fill in the song *In The Air Tonight*?

I stress that such mnemonic processes of making beliefs propositionally available is not the same as making dispositional beliefs *occurrent*.[64] Dispositional beliefs become occurrent when their manifestation conditions are met, e.g. when you are being asked a question about your belief. Additionally, dispositional beliefs influence behavior: having the dispositional belief that p entails *behaving* like you believe that p (Schwitzgebel,

[64] Nor can it be said that we had the 'disposition *to* believe' the proposition that was made available *for* belief by the episodic memory experience (but it *can* be said that we have the disposition to believe the proposition *after* the episodic memory; this is precisely what I mean with 'making a proposition available *for* belief'); c.f. Audi (1994).

2011). Neither is the case for the above-mentioned 'belief' that was not propositionally available. As the examples above show, the 'belief' that there is an elephant in the room, or the 'belief' that your friend hit the drum six times, neither became occurrent when the subject was asked about them, nor did they directly influence behavior.

So much for the problem of novelty. Next: the problem of reliability.

The reliability of our declarative memory is a controversial and thorny issue; see e.g. Loftus and Pickrell (1995); Loftus (1997); Gardner (2001); Senor (2019); Frise (2021, 2022). Memory is not an infallible and unmalleable preserver of past beliefs and perceptions. On the contrary, it is well-known that, for example, what we believe to be our episodic memories are strongly formed, re-shaped and influenced by our background beliefs, primes, hopes, expectations and many other subject-dependent and context-dependent factors — at the moment of memory-creation, at the moment of memory-retrieval, and at many moments in between. Often what we believe to be our episodic memories are not actually *memories* (which require that the imagined event actually happened and that there is an appropriate causal connection from the remembered event to the rememberer's current mental state), but rather 'mere imaginings'. Thus the question arises to what extent (what we believe to be) our memories are *reliable* and *robust* sources of true beliefs: if we believe that we remember some event, then under which conditions do we reliably and robustly *truly believe* that it did in fact happen?

The reliability of episodic memory is dependent on many conditions of many different types — too many to mention here. All that is relevant for my present purpose is that episodic memory *can* be reliable and robust. I think everybody (save some rare exceptions, e.g. Frise (2022)) would find this acceptable. This is all I need because I shall argue for the conditional claim that imagination is reliable *only insofar* as the memories that it draws on are reliable and robust. I argue only for this conditional claim. If one believes that memory is rarely or never reliable or robust (a claim to which I do not subscribe but will not argue against), then this would

imply that imagination is also never reliable or robust.

As a final note, I emphasise that novel beliefs gained through memory, such as the ones mentioned above, are *justified* iff the episodic memory that sourced the novel belief is reliable and robust, per (3.4), which I assumed *can* be the case. This justification need not be based on other (already justified) beliefs. Hence, it seems that memory can *generate* not only novel beliefs but also novel *justification*.[65] Consequently, it seems that memory can generate full-fledged *knowledge* entirely 'on its own' — that memory is a so-called *basic source* of knowledge (Section 3.5.1). This is not the case. The above-mentioned examples are simplified and, as such, somewhat deceptive. As I will argue in the next Section, once we look at the process of obtaining novel beliefs through memory carefully, we recognise that memory cannot yield novel beliefs entirely 'on its own' because there are always auxiliary *beliefs* involved in this process — beliefs *about* the accuracy of our memories.

To recapitulate, I have discussed different types of memory, and I specified that the type of memory most relevant for the current Chapter is *episodic imagination* (3.5). I next discussed several epistemological problems haunting episodic memory: the problem of *novelty* (i.e. that memory cannot yield *novel* beliefs) and the problem of *reliability* (i.e. that memory cannot reliably yield *true* beliefs). I argued against the problem of novelty by providing counter-examples, and I concluded that memory can be a source of genuinely *novel* beliefs. I then noted that the problem of reliability is a serious problem haunting imagination, and that this problem is too big to handle in this current Chapter. But I also mentioned that nearly *all* philosophers thinking about memory agree that memory *can be* reliable. This is all I need for my present purpose: I shall argue in this Chapter only for the conditional claim that *imagination is reliable insofar as memory is reliable.*

Enough about memory. I turn to the concept of *quasi-perception.*

[65] Memory can of course also *preserve* justification, e.g. if we forgot, and then remember, a previously justified belief (and the reason for its justification).

3.3 Quasi-perception

3.3.1 Perception and quasi-perception

In this Section, I explain the concept of quasi-perception and its relation to ordinary perception.[66] I repeat that ordinary perception is the inclusive disjunction of ordinary vision (2.1) and its other sensory modality counterparts (hearing, etc.); recall Chapter 2, Section 2.2, p.24. On the details of the content of ordinary perception I elaborate below. I shall mainly build on the discussion of quasi-perception by Sartre (1948), Huemer (2001), Gendler (2004) and Nanay (2015).

There are *two* types of mental states that I call *quasi-perceptual*: mental states of action-imagination (3.2), where the imagined action is imagined perception, and mental states of episodic memory (3.5). Explication:

> [Quasi-perception] Subject S quasi-perceives entity ε iff S imagines perceiving ε (3.2), or S has an episodic memory about ε (3.5), or both. (3.7)

Thus there are two types of *processes* that I call *quasi-perceptual processes*: acts of imagination (3.1) and episodic memories (3.5). Acts of imagination and episodic memories are both quasi-perceptual and hence share important similarities, but there are also important differences between the two, as I shall describe below.

I begin by noting three important *similarities* between ordinary perception and quasi-perception, treating *remembered* and *imagined* quasi-perception as one and the same.[67] I shall turn to the distinction between remembered and imagined quasi-perception when I discuss the differences between ordinary perception and quasi-perception.

---

[66] I shall henceforth call perception *ordinary perception* to highlight the difference with quasi-perception.

[67] I here partly follow (Huemer, 2001, Ch.IV), who argues that the three essential characteristics of perception are (i) sensory qualia are involved, (ii) they give rise to mental states with representational content, and (iii) they are *forceful* (see below).

Similarities between ordinary perception and quasi-perception

The first similarity is that ordinary perception and quasi-perception both give rise to mental states with *representational content*: both processes involve mental states with sensory or motory content that stands in a representation-relation to an entity or class of entities.[68] This is neatly accounted for by my explications of ordinary perception (2.1) and action-imagination (3.2), as representing sensory or motory content is a necessary condition for both. This representation-relation is crucial, epistemically speaking: if the represented entity or class of entities is part of the natural world, then it is precisely in virtue of this representation-relation — in fact, *only* in virtue of it — that quasi-perception can *concern* the natural world at all, hence possibly enable us to *learn* about the natural world, even though the quasi-perceptual content is not (directly) *caused* by the natural world. More on this below.

The second similarity between ordinary perception and quasi-perception concerns the *determinacy* of properties of the (quasi-)perceived scenario. This similarity is rather intricate, and somewhat controversial, so I will spend some time on it.

Hume (1896) (in)famously distinguished between ordinary perception, memory and imagination on the basis of the *vivacity* (or 'liveliness') of the (quasi-)perceived scenario: according to Hume, ordinary perceptions have the highest degree of vivacity, remembered quasi-perceptions have a lower degree of vivacity and imagined quasi-perceptions have an even lower degree of vivacity still — they are the 'faintest copies' of ordinary perceptions. Hume (1896) wrote:[69]

> [T]he ideas of the memory are much more lively and strong than

[68] This similarity implies that both ordinary perception and quasi-perception are types of sensory *experiences* and, hence, *sensory qualia* are involved in both. I acknowledge that in the past it has been explicitly argued that there are no sensory qualities involved in quasi-perception, but current consensus seems to be that they *are* involved; c.f. Noordhof (2002) and the references therein for nuanced arguments in favor of and against the idea that there are sensory qualia involved in quasi-perception. [69] As quoted and discussed in (Huemer, 2001, p.78).

> those of the imagination, and [...] the former faculty paints its objects in more distinct colours, than any which are employed by the latter. [...] [I]n the imagination the perception is faint and languid, and cannot without difficulty be preserved by the mind steady and uniform for any considerable time.

Although Hume's distinction between perception, memory and imagination on the basis of their respective *vivacity* is plausible at first sight, current consensus is that it is misguided: it is simply false, phenomenologically speaking, that for all human beings remembered quasi-perceptions are always more vivid than imagined quasi-perceptions, and it is certainly false that ordinary perceptions are always more vivid than *quasi*-perceptions. In any case, the concept of vivacity is rather vague and out of vogue, and spending time explaining it would take me too far off track; c.f. Govier (1972); Dauer (1999); Owen (2008). I lay it aside here and turn towards a more contemporary way of describing the similarity and difference between the content of ordinary perception, memory and imagination.

A promising and fashionable way of describing differences in perceptual content along roughly the same lines as Hume's *vivacity*-distinction is in terms of the *determinacy of properties* of a (quasi-)perceived scenario. Properties can be *determinate* to various degrees; some properties can be more determinate than others. The relative relation between the determinacy of properties is known as the *determinable–determinate relation* (Wilson, 2023). Determinable properties are properties that can be made more specific; determinate properties are properties that have been made specific. To give one example: shape is a determinable property; rectangular is a determinate property relative to the determinable property shape; and square is a so-called *super*-determinate property relative to the determinable property rectangular (and, transitively, to the property shape), meaning that it is not a determinable property, it cannot be made *more* determinate (relative to the determinables rectangular and shape).

Determinable properties are *made* determinate: we make properties determinate by perceptually paying *attention* to the relevant features of

the perceived scenario. If a given perceived shape is rectangular, for example, then we can make this property more determinate by paying attention to the length of its sides: if we perceive all sides to be of equal length, then the shape is a square; if we instead perceive some sides to be of unequal length, then the shape is a proper (non-square) rectangle.

With the determinate–determinable relation of properties in hand, I introduce the popular account of ordinary perceptual content from Nanay (2010, 2015):[70]

> Our perceptual apparatus attributes various properties to various parts of the perceived scene. [...] Perceptual content is constituted by [the sum-total of] the properties that are perceptually attributed to the perceived scene. [...]
>
> Some of the properties we perceptually attribute to the perceived scene are determinates or even super-determinates. Some others, on the other hand, are determinable properties. [...]
>
> [P]erceptual attention should be thought of as a necessary feature of perceptual content (Nanay, 2010, 2011). More precisely, attention makes (or attempts to make) the attended property more determinate [...]. If I am attending to the color of my office telephone, I attribute very determinate (arguably super-determinate) properties to it. If, as it is more often the case, I am not attending to the color of my office telephone, I attribute only determinable properties to it (of, say, being light-colored or maybe just being colored). In short, attention makes (or attempts to make) the perceived property more determinate.

So, for Nanay, perceptual content is the sum-total of all determinable and (super-)determinate properties that we attribute to the perceived scenario; and determinable properties of a perceived scenario are made (or are attempted to be made) determinate by paying *attention* to them.

Turning, then, to the content of *quasi*-perception — mental imagery — Nanay (2015, p.1728–9) continues:

> I outlined a simple, and not particularly controversial, account of

[70] Quoted from (Nanay, 2015, 1727–8).

> perceptual content in the last section. But what is the content of mental imagery? My answer is that the content of mental imagery is exactly the same as the content of perceptual states.
>
> More precisely, our imagery attributes various properties to various parts of the imagined [or remembered] scene. The content of imagery is the sum total of the properties attributed to the imagined scene. Some of these properties are determinates or even superdeterminates. Some others are determinables. Attention makes (or tries to make) the attended property more determinate.

So, Nanay holds that the *type* of content of quasi-perception is exactly the same as the type of content of ordinary perception: they are both just the sum-total of determinable and determined properties attributed to the (quasi-)perceived scene. I note that my explications of ordinary vision (2.1) and 'imagistic' imagination, e.g. action-imagination (2.30), harmonise perfectly with Nanay's account. So, let us accept this as the second similarity between ordinary perception and quasi-perception.

At this point, I wish to emphasise one feature of the content of quasi-perception — mental imagery — about which there is mild consensus that is it also present in ordinary perceptual content: *spatial properties*, or at least a functional analog thereof; recall the Conceptual Basis for this Thesis in Chapter 2, Section 2.2; c.f. Kosslyn and Pomerantz (1977); Kosslyn (1980); Nanay (2021). In the Conceptual Basis, I mentioned that mental imagery has besides semantic properties (content, reference, etc.) also *spatial properties*, or at least a functional analog thereof, e.g. in the sense that spatial distances between parts of the mental image are defined "in terms of the number of discrete computational steps required to combine stored information about them" (Pitt, 2022, §5). As I said above, it is mildly controversial whether ordinary perceptual content also has spatial properties (I neither assumed nor denied this in my Conceptual Basis; Chapter 2, Section 2.2), but it is regularly argued that it does (Macpherson and Bermudez, 1998; Thompson, 2010) — although there is an ongoing debate about 'on which level' of perceptual content spatial

content would reside; c.f. (Pacherie, 2000; Kulvicki, 2007). If it does, then this marks the third important similarity between perceptual content and quasi-perceptual content; if it does not, then this marks the first important *difference* between the two. (In any case, it is important to note that the content of quasi-perceptual content uncontroversially *does* have spatial properties — this will be relevant in Sections 3.4 and 3.5.)

Differences between ordinary perception and quasi-perception

Now the question arises: what, then, are the clear *differences* between ordinary perception and quasi-perception, since there undeniably *are* clear differences? Following the quote above, Nanay (*ibid.*) continues:

> The only difference concerns where the extra determinacy comes from. As we have seen, both in the case of perceptual content and in the case of mental imagery, attention makes the attended property more determinate. This increase in determinacy in the case of perception comes from the sensory stimulation: if I am attending to the color of the curtain in the top left window of the building in front of me, this color will be more determinate than it was when I was not attending to it. This difference in determinacy is provided by the world itself — I can just look: the exact shade of the curtain's color is there in front of me to be seen.
>
> In the case of mental imagery, this difference in determinacy, in contrast, is not provided by the sensory stimulation, for the simple reason that there is no sensory stimulation that would correspond to what I visualise: if I visualise the house I grew up in and you ask me to tell what exact color the curtain in the top left window was, I can shift my attention to that color and I can even visualise the exact color of the curtain. However, this increase in determinacy is not provided by the sensory stimulation (as I don't have any), but by my memories (or what I take to be my memories) or my beliefs or expectations.

Nanay argued that the difference between perceptual content and quasi-perceptual content lies not in the *character* of the content but rather in

the *source* of its determinacy. In the previous Chapter, I denoted this difference by saying that ordinary perception is *exogenous*, while imagination *endogenous*. But I nonetheless wanted to introduce Nanay's account of this difference, because it enables us to look at the differences between exogenous and endogenous mental states in a bit more detail, as follows.

To begin, I note that Nanay described the *main* source of determinacy of perceptual content, since determinacy in ordinary perception can also come from memories, background beliefs, expectations, desires, hopes, etc., just like it can for quasi-perception.[71]

Having said this, I acknowledge that the (exogenous) content of ordinary perception is never *directly* provided by the natural world, in the sense that the perceiving subject necessarily plays a crucial constitutive role in determining perceptual content (viz. the theory-ladenness of observation, etc.); see Section 3.3.2 for more discussion. Notwithstanding, it can reasonably be said that the content of ordinary perception is *much more directly* provided by the world than the content of quasi-perception is. With this caveat in place, for the sake of convenience I henceforth just say that the content of ordinary perception is *directly* provided by the natural world.

So, let us accept that the (main) source of determinacy of properties of a perceived scenario is, directly, the world. This, then, leads us to an important difference with quasi-perception, because the main source of determinacy in quasi-perception is *not* the natural world directly. The content of quasi-perceptual content — mental imagery — is endogenous, not exogenous. What, then, *is* the (endogenous) source of determinacy in quasi-perception? Here the difference between episodic memories and acts of imagination becomes important. For the sake of clarity, I next contrast 'pure' memory (i.e. a mental state of memory without imagined content) with 'pure' imagination (i.e. a mental state of imagination with-

[71] If the world itself is not the main source of determinacy of perception, then perception will generally be *falsidical*, hence *unreliable*. It is tempting to propose the following biconditionals: perception is veridical iff the source of its determinacy is the natural world; and perception is reliable iff the *main* source of its determinacy is the natural world.

out mnemonic content).

Like the exogenous content of ordinary perception, the content of 'pure' memory is also provided by the natural world. But this time it is not provided *directly* by the natural world but instead *indirectly* through some "appropriate causal chain"; recall Section 2.6.3. This causal chain involves preservative mnemonic processes in the brain and body of the quasi-perceiver, and it does not involve direct sensory input, hence we can say that the content of 'pure' memory is endogenous rather than exogenous, but also that it is provided *indirectly* by the natural world.

The endogenous content of 'pure' imagination, by contrast, is not directly or indirectly determined by the world at all but *only* by subject-dependent factors: voluntary choices, expectations, background beliefs, *aliefs*, desires, hopes, emotions, moods, etc.; recall Section 2.8.1. As I have acknowledged multiple times, Hume's *recombination principle* implies that the content of 'pure' imagination is *also* indirectly provided by the natural world. I believe this is true, but I repeat my comment from footnote 48 (p.75): this imagined content is even *much more indirectly* provided by the natural world, to the extent that we can reasonably say that, compared to memories, for all practical purposes, this content is *not* indirectly provided by the natural world — it is not only endogenous but also *voluntary*.

Presenting the differences in content in this way makes things deceptively simple. As I have said and argued many times, the distinction between memory and imagination is vague, not sharp. In the overwhelming majority of cases, a mental state of imagination will not be 'pure' but will instead be thoroughly infused with mnemonic content. This is what Nanay means when he writes that the source of determinacy in the content of imagination is provided by, amongst many other factors, memories.

Now, contrary to what Nanay seems to suggest, I do believe that this difference in source of determinacy entails a difference in content along the lines that Hume alluded to with his concept of vivacity. I believe so for the simple reason that the source of determinacy in quasi-perception

— *ourselves* — is much more easily exhausted than the world. Surely, we should expect that a quasi-perceived scenario has much *fewer* determinate properties than a perceived scenario — if this difference is not *necessary*, it surely is *typical*. In the famous words of Sartre (1948, p.9):

> In the world of [ordinary] perception, no 'thing' can appear without maintaining an infinity of relations to other things. Better, it is this infinity of relations — as well as the infinity of the relations that its elements support between them — it is this infinity of relations that constitutes the very essence of a thing. Hence a kind of *overflowing* in the world of 'things': there is, at every moment, always infinitely more than we can see; to exhaust the richness of my current perception would take an infinite time. [...]
>
> But in the [quasi-perceived] image, on the other hand, there is a kind of essential poverty. The different elements of an image maintain no relations with the rest of the world and maintain only two or three relations between themselves: those, for example, that I could note, or those that it is presently important to retain. It should not be said that the other relations exist in secret, that they wait until a beam of light moves on them. No: they do not exist at all.

Similarly, in the more recent words of Huemer (2001, p.77):

> One difference [between perception and quasi-perception] would typically be that the perceptual experience has a more specific and detailed content; [...] [for example:] imagine a newspaper. There is no difficulty in doing this. However, your "image" of a newspaper in this case is not as detailed as a visual experience of a newspaper. If you are having a visual experience of a newspaper, you can thereby read said newspaper. But I doubt you will find yourself able to read the newspaper you are merely imagining. The mental image is too indeterminate in its representational content.

I repeat, there is nothing *necessarily* stopping you from having an incredibly detailed mental image of a newspaper. Someone with hyperthymesia or photographic (eidetic) memory may be able to genuinely *read* this morning's newspaper in their quasi-perceptual memory of it. But typically this

will not be the case: as sources of determinacy, our memories and other subject-dependent factors are typically easily depleted, whereas the natural world is not. In other words: in quasi-perception there is typically less that *can be paid attention to* than in ordinary perception. But, contrary to what Sartre argued above, even if there is much *less* that can be made determinate in a quasi-perceived scenario, this does not mean that there is *nothing* that can be made determinate in quasi-perception: we can make features of our *memories* more determinate *in the imagination*; recall my discussion of memory as a source of novel beliefs in Section 3.2.3, and see Section 3.5.2 below.

I turn to the second difference between perception, memory and imagination that I wish to highlight. This difference is purely phenomenological. It is widely accepted that ordinary perceptual experiences have a distinctive phenomenology: they distinctively present their content as *actual*; that is, as being *right there, like that, in the world*. This appearance is often described in terms of phenomenal "seeming": the content of perception *seems* actual to us (Brewer, 1999; Huemer, 2001; Markie, 2005; Chudnoff, 2011; Chudnoff and Didomenico, 2015). The content of quasi-perception, by contrast, does *not* seem actual to us. The content of action-imagination, explication (3.2) tells us, seems *possible* (and is *accepted* as such), not actual. The content of (what we believe to be) episodic memory seems *like it happened in the past*. Huemer (2001, p.77–78) writes:

> Even if you have a very vivid, very detailed imagination, or if you have very poor eyesight, you still would never confuse seeing a tomato with imagining one. [...] The reason lies in what I call the "forcefulness" of [ordinary] perceptual experiences: perceptual experiences represent their contents as actualized; states of merely imagining do not. When you have a visual experience of a tomato, it thereby seems to you as if a tomato is actually present, then and there. When you merely imagine a tomato, it does not thereby seem to you as if a tomato is actually present. [...]
>
> A memory experience is distinguished from a perceptual experience chiefly by the fact that the object of a memory experience seems

	Source of content	**Phenomenology**
Perception	The natural world (directly)	Content seems actual
Memory	The natural world (via an appropriate causal chain)	Content seems past
Imagination	Choice, background beliefs, aliefs, expectations, desires, hopes, moods, emotions, etc.	Content seems possible

Table 3.1: Differences between perception, memory and imagination.

> to the subject to be something that happened in the past, and this has nothing to do with how colorful or faint the memory may be.

So much for the similarities and differences between the content and phenomenology of ordinary perception, episodic memories and acts of imagination. See Table 3.1 for an overview.[72] I next explain in more fine-grained detail how perceptual and quasi-perceptual processes yield (quasi-)perceptual *beliefs*.

3.3.2 Perceptual belief and quasi-perceptual belief

As I said in the previous Section, the two types of mental processes that I regard *quasi-perceptual* are episodic memories and acts of imagination. These two quasi-perceptual processes can yield quasi-perceptual beliefs. As Gendler (2004, p.1152) put it: beliefs are *quasi-perceptual* when "the presence of a mental image [plays] a crucial cognitive role in the formation of the belief" — similarly to how ordinary perceptual content plays a 'crucial cognitive role' in the formation of ordinary perceptual beliefs, I add. In this Section I make this more precise. To provide a clear contrast, I begin with describing how we obtain ordinary perceptual beliefs.

I shall follow Dorsch (2016b) and hold that a belief is a *perceptual* belief iff the belief is rationally determined by an ordinary perceptual process.

[72] It seems plausible that Hume's concept of *vivacity* was meant to capture *both* the difference in source of (determinacy of) content and the difference in phenomenology displayed in this Table.

On the "rational determination" of ordinary perceptual beliefs, Dorsch (2016b, p.90) explains:

> [T]here are two distinct aspects of perceptual belief that are in need of rational determination by the underlying perceptual experiences.
>
> First, which *content* the belief has — that is, which proposition is endorsed — has to be a matter of how the experience presents things in our environment as being. [...] For example, we come to believe (and know) *that it rains* because we experience *the rain*. [...]
>
> Second, which *attitude* we adopt toward the propositional content in question (i.e. the attitude of belief) has to be a function of what kind of experience is concerned (i.e. perceptual experience). We come to *believe* (and know) that it rains because we *see* the rain.

So, according to Dorsch, the rational determination (henceforth just: determination) of ordinary perceptual beliefs can and should be epistemologically reconstructed as a process that involves *two distinct steps*. In terms of my explication of perception (2.1), this two-step process can be formulated as follows

The two-step schema for ordinary perceptual beliefs:
(1) On the basis of perceiving (concrete observable) entity ε, proposition p with topic ε comes to mind; (3.8)
(2) On the basis of this process being a process of *perception*, the propositional attitude of *belief* is adopted to p.

I next make two comments about this two-step process (3.8) that will be relevant for what is to come.

Firstly, step (1) states that on the basis of an ordinary perceptual process, a relevant proposition comes to mind. I have described this process at length in the previous Section, focusing on how properties of a perceived scenario are determined by paying attention to relevant features of that perceived scenario. Here I wish to add some additional comments on this process that are particularly relevant for perceptual (propositional) *beliefs*.

Following any given perception, *many different* propositions may come to mind, and *which* particular proposition comes to the mind of a given subject S on the basis of a given perception depends not only on the objective details of the perceived scenario, but also depends crucially on *subjective* factors, notably on the background knowledge and readily-available perceptual concepts of S (viz. the 'theory-ladenness' of perception), on the salience of features of the perceived scenario and on what S pays *attention* to, i.e. on the degree to which properties of the perceived scenario are determinate (Nanay, 2010, 2011), on the expectations and epistemic purpose of S, and on the primes, moods, emotions, desires and hopes of S and many other subjective and contextual factors (Dorsch, 2016a, fn.3).

The important consequence of this crucial role of subjective factors in determining which proposition comes to mind is that perceiving the *same* scenario twice can make *different* propositions come to mind if the just-mentioned subjective factors have changed enough in between the 'same' perceptions. The first two subjective factors mentioned above are particularly important for the purpose of this Chapter, so let me give an example that concerns these two factors similar to an example that I presented before. Suppose that you perceive an elephant in front of you and thus the proposition comes to mind *that* there is an elephant in front of you. Then, a friend comes by who reminds you that African elephants generally have much larger ears than Asian elephants. You now look at the elephant again and pay particular attention to the size of its ears, which seem to you as rather small for an elephant. Now the proposition comes to mind that there is an *Asian* elephant in front of you. The difference between the two propositions in this example may be small, but they are distinct propositions nonetheless; and more radical examples are easily conceived.

Secondly, step (2) states that on the basis of this process being a process of *ordinary perception*, we adopt the propositional attitude of *belief* to the proposition that came to mind. As I said in the previous Section (recall Table 3.1), it is widely accepted that ordinary perceptions have

a distinctive phenomenology: it *seems* to us as being right there, like that, in the world. We believe propositions about the content of our ordinary perceptions *because* these propositions come to mind on the basis of *perceptions*, rather than e.g. on the basis of an act of imagination. Ordinary perceptions and perceptual beliefs generally come hand-in-hand, often involuntarily so.

In fact, ordinary perception yields perceptual beliefs so 'automatically' that it is easy to forget that this second step in Dorsch's (2016b) two-step process of determining perceptual beliefs is even there. Indeed, we are generally *aware* of this second step only in those cases where we hesitate to adopt the attitude of belief to the proposition that presents itself, which is the case, for example, when the conditions under which we perceive are clearly suspicious (e.g. there is bad light, our eye-sight is compromised, or the perceived object is an ostensible optical illusion) or the belief itself is extra-ordinary (e.g. when you perceive an elephant in the room, which is rather strange, you may rub your eyes and look again, or ask your neighbour to come confirm that there is, indeed, an elephant in the room). Notwithstanding, Dorsch (2016b, §IV, my italics) insists:

> What is crucial here — and sometimes overlooked — is the fact that the rational determination of the *content* of propositional knowledge happens independently of the rational determination of its belief *attitude*.

As the last-mentioned example above shows, these two steps may even involve different mental states at different moments in time: you perceive *that* there is an elephant in the room at some moment, but you only adopt the attitude of belief to this proposition at a much later moment, e.g. once your neighbour confirms that there is indeed an elephant in the room.

I flag that the second step in the two-step schema for ordinary perceptual beliefs (3.8) — the rational determination of the attitude of belief — does not amount to full-fledged *epistemic justification* of a perceptual belief (3.4). A popular way of phrasing what happens in this second step is that perceptual beliefs are *prima facie justified*: due to the distinctive

phenomenology of ordinary perception, perceptual beliefs present themselves to us "forcefully" (Huemer, 2001) and are to some extent *directly* justified in virtue of being *perceptual* beliefs — i.e. perceptual beliefs justify *themselves* to some extent, because they are not justified in virtue of some *other*, already justified, belief. I however avoid using the notion of *prima facie justification* in this Thesis, as it is a Pandora's box that I prefer to keep closed; c.f. Senor (1996); Markie (2005); Goldman (2008); Chudnoff (2011); Hasan and Fumerton (2022).

Moving on to quasi-perception. I begin by noting that I, like Dorsch, emphasise the second step in the two-step schema for ordinary perceptual beliefs (3.8) so much because herein lies the crucial difference between the formation of ordinary perceptual beliefs and the formation of quasi-perceptual beliefs. Quasi-perceptual processes — episodic memories and acts of imagination — also yield beliefs in two steps, and only the second step is markedly different than it was for the case of ordinary perceptual beliefs (3.8). It is not the case that, on the basis of a quasi-perceptual process being a process of *imagination*, we adopt the propositional attitude of belief to a proposition that comes to mind. Rather the contrary: we generally do *not* adopt the attitude of belief to propositions that come to mind on the basis of an act of imagination, *because* it is a process of *imagination*. Imagination is "epistemically innocent", as Balcerak Jackson (2016, p. 44) called it: imaginings, on their own, cannot support beliefs about the natural world. Kind (2013, p.6) wrote that imagination is not 'world-sensitive': changes in our environment correspond to changes in our perception and in our (perceptual) beliefs, but they need not, and often do not, correspond to changes in the content of our imagination. Additionally, in the previous Section, I explained how the content of imagination does not *seem* to us as being there, like that, in the world, but rather only *seems possible*. So, imagination cannot yield beliefs about the natural world on its own. Something extra is required, something *external to* imagination: we need a new second step for the two-step schema for quasi-perceptual beliefs.

What *could* motivate us to adopt the attitude of belief to propositions *about the world* that come to mind on the basis of a 'mere' quasi-perceptual experience? Reasonably, Dorsch (2016b) suggests the following:[73]

The two-step schema for quasi-perceptual beliefs:
(1) On the basis of quasi-perceiving (concrete observable) entity ε, proposition q with topic ε comes to mind;
(2) On the basis of the *meta-belief* that the quasi-perceived scenario accurately represents the natural world, the propositional attitude of *belief* is adopted to q. (3.9)

Here, the first step in (3.9) remains (structurally) the same as for ordinary perceptual beliefs: directly on the basis of a quasi-perceptual process, i.e. by having an episodic memory or by performing an act of imagination, relevant propositions may come to mind. This is the sense in which Gendler means that "the presence of a mental image plays a *crucial* cognitive role in the formation of the belief", recall the quote at the beginning of this Section: if there were no quasi-perceptual process, then proposition q would not have come to mind. But — this is the important point — although a quasi-perceptual process plays a *crucial* role in the formation of a quasi-perceptual belief (it is responsible for the *first* step in the two-step process), it does not play *every* role in the formation of a quasi-perceptual belief (it is not responsible for the *second* step in the two-step process).

The second step in this two-step schema for quasi-perceptual beliefs (3.9) is markedly different from the second step in the two-step schema for ordinary perceptual beliefs (3.8). *Quasi-perception does not and cannot, on its own, motivate us to adopt the attitude of belief to propositions about the natural world.* Therefore, quasi-perceptual beliefs must always

[73] In this Chapter, I limit my attention to quasi-perceptual beliefs about in principe *observable* matters of fact. Although he does not say it explicitly, Dorsch (2016b) seems to make the same assumption. For quasi-perceptual processes that concern e.g. non-existent or even *impossible* entities and scenarios, see next Chapter.

be motivated by something *external* to the quasi-perceptual process: they must be motivated by ancillary *meta*-beliefs[74] that the content of the quasi-perceived scenario *accurately represents* the natural world. Dorsch (2016b, p.102) writes:

> The only concession to be made is that imaginative experience can play only part of the [crucial] grounding role of perceptual experience, namely the part concerned with the determination of content. [...] [T]he attitude of the resulting belief is, by contrast, determined by our [meta-]beliefs about the accuracy of how our imaginative experience visually presents things as being.

In Section 3.3.4, I shall argue in more detail why this is the case. To pave the way for this argument, in the next Section I present three examples of acts of imagination that yield quasi-perceptual beliefs.

I note, finally, that the discussion thusfar concerns quasi-perceptual *beliefs*, not quasi-perceptual *knowledge*. Quasi-perceptual beliefs do not automatically amount to knowledge: the quasi-perceptual belief may still be false or unjustified, even though it is 'rationally determined' per the two-step schema (3.9). Sources of quasi-perceptual beliefs are, of course, not *infallible* sources of knowledge — quite the contrary. To quasi-perceptual *knowledge* I shall turn in Section 3.2.2.

3.3.3 Three examples

I next present three examples where quasi-perceptual beliefs are obtained.

Example 1.[75] Think about your bedroom at home (if you're cur-

[74] What I call a *meta-belief* in step 2, Dorsch (2016b) calls an *ancillary belief*. I prefer to call it a *meta-belief* because the "meta" indicates more clearly what this ancillary belief is *about*: it is about the *relation* between the quasi-perceived scene and the natural world. As a side-note, I admit that I find it tempting to formulate the second step in the two-step determination of ordinary perceptual beliefs (3.8) *also* as: "(2) On the basis of the belief that the perception accurately (re)presents the actual world, the propositional attitude of *belief* is adopted to that proposition". It is highly contentious, however, whether perceptual beliefs *are* necessarily grounded in other beliefs; e.g. the (dispositional) belief that one's perception is *veridical*. Although I personally believe that this is the case, arguing for this would take me too far off track, so I keep the two two-step processes, (3.8) and (3.9), distinct. [75] This example comes from (Gendler, 2004, §3).

rently in it: close your eyes) and ask yourself the following question: if you removed all the furniture from your bedroom, could an elephant fit comfortably inside? And what about two elephants, or three, or four? What is the maximum amount of elephants that would fit in your bedroom? Suppose you find yourself imagining that you could comfortably fit two elephants in your bedroom, but that three elephants would be much too tight a squeeze (α). Upon a brief moment of reflection, you convince yourself that your imagined bedroom and imagined elephants are the correct size, or at least close enough to it, so you adopt the attitude of belief to α.

Example 2.[76] Suppose that you wrote a poem for your best friend's wedding and you want to fit this poem to Queen's *We Will Rock You*. You have the poem written on paper and want to check whether its length and rhythm matches that of the chorus of the song. Because you currently find yourself in a library — *silence!* — and you cannot play the song out loud and sing your poem along, you *imagine* playing the song in your head and *imagine* singing the words of the poem along with the song to see whether they match up. You find that the poem is finished before the song is. The proposition comes to mind that the poem, when sung out loud, does not fit to Queen's *We Will Rock You* (β). You try again and find the same result. You convince yourself that you sang *We Will Rock You* accurately, and so you adopt the attitude of belief to β.

Example 3.[77] Imagine the house that you grew up in and ask yourself the question: how many windows does it have? (Do it!) If you did not have the answer immediately available in your memory, then presumably you tried to find an answer by creating a mental representation of each room (based on your episodic memories of them)

[76] This example comes from (Dorsch, 2016b, p.95). [77] This example comes from Nersessian (2018, p.309), who attributes its origin to the psychologist Herbert Simon. Elsewhere (on the internet) the origin of this example is occasionally attributed to the psychologist Alan Baddeley.

> and then *counting* the windows in your imagination. The proposition then came to your mind that the house you grew up in has, say, n windows (γ). After a quick re-count, you convinced yourself that you did not forget to count any windows and that your memories are reliable, so you confidently adopt the attitude of belief to γ.

Each of these three examples are epistemic acts of imagination that yield beliefs about the natural world. These beliefs are quasi-perceptual beliefs because the belief-forming process in all three examples satisfy the two-step schema of quasi-perceptual beliefs (3.9): *first*, on the basis of some quasi-perceptual process — imagining, respectively, fitting elephants in your bedroom (example 1), singing a poem to a song (example 2), seeing and counting windows (example 3) — propositions came to mind; *second*, on the basis of meta-beliefs about the *accuracy* of your imaginings, the attitude of belief was adopted to the propositions that came to mind.

There are also important differences between each example. Example 3, for instance, notably involves not only imagination but also a hefty dose of *memory*. As a result, the quasi-perception in example 3 brings to mind a novel proposition, not through some purely imaginative recombination of ideas, but rather by *drawing attention to features of our memories that we did not pay attention to before*. In other words: example 3 is an example where we use imagination to use episodic memory to make beliefs propositionally available; recall Section 3.2.3. Examples 1 and 2, by contrast, are more 'creative' and less mnemonic, as they concern quasi-perceptual beliefs that you could not obtain through any pure memory process, simply because you have never contemplated or experienced the imagined scenarios of these examples.

Another important difference between these three examples is that different *types* of quasi-perceptual content play an important epistemic role in each. In Example 1, *spatial properties* are epistemically relevant: the relative sizes of elephants and the room, and the ways in which they do or do not fit together, are consequences of the *spatial* properties of your imagined elephants and your imagined room. Example 2 is similar in this

regard, although it concerns the *auditory* analogue of spatial properties (e.g. the *duration* of a song). Example 3 is again different from Examples 1 and 2 in this regard, as arguably spatial properties of the quasi-perceived scenario are not epistemically relevant in Example 3: it is only the *number* of windows that matters crucially, not the sizes of the windows, or their positions and relative distances, etc. (Although, intuitively, imagining the spatial properties of these scenarios correctly will improve the odds of the quasi-perceptual beliefs obtained in Example 3 being *reliable* and *robust*, i.e. justified.)

I next make some elucidatory remarks about the quasi-perceptual beliefs obtained in these examples.

(a) It may be argued that the 'real' source of belief in these examples is not imagination but *memory*: these examples are examples of predominantly *mnemonic* processes, not of 'pure' acts of imagination. This objection is most applicable to example 3: here, we have a quasi-perception of a *remembered* scenario (the rooms with windows in the house you grew up in), so if this process yields a belief then the source of this belief surely is memory. I respond that this example shows exactly the difficulty in disentangling the contribution of memory from the contribution of imagination to processes of quasi-perceptual belief acquisition. In fact, these three examples were deliberately chosen to demonstrate an increasing amount of mnemonic content present in the act of imagination — to show the intertwinement of memory and imagination *and* to show their independence. Having said this, I insist that example 3 is an example where imagination too — not only memory — is crucially involved in the production of a quasi-perceptual belief. The mental simulation that you performed while counting the windows did not involve *only* mnemonic content: no, you counted *remembered* windows in your *imagination*. So, if anything, example 3 shows that memory and imagination can *co-operate* to yield quasi-perceptual beliefs in distinctive and epistemologically interesting ways — I shall call such beliefs *distinctively imaginative beliefs*; see Section 3.5.2.

(b) It may be argued that these examples (particularly examples 1

and 2) are not cases where imagination is a source of quasi-perceptual beliefs but rather cases where imagination is a source of *modal* beliefs, i.e. beliefs about *(im)possibilities*. The belief obtained in example 1, for instance, can be formulated as: Two elephants would fit in my bedroom but three elephants would not, or, equivalently, It is (practically) possible to fit two, but not three, elephants in my bedroom. This surely is a modal belief. To respond to this objection, I first note that quasi-perception is *crucially* involved in the process of obtaining this belief, in the sense that the proposition would not come to mind if there were no quasi-perceptual process. So, if anything, this example shows that processes of obtaining modal beliefs and quasi-perceptual beliefs are *compatible* with each other rather than mutually exclusive; recall Section 3.2.2. Secondly, I note that a modal belief such as It is (practically) possible to fit two, but not three, elephants in my bedroom is nearly equivalent in semantic content to the belief that The size of my bedroom is larger than two, but smaller than three, elephants combined, which *is* a non-modal belief about contingent matters of fact about the actual world. Both beliefs (the modal and non-modal one) can be obtained through the quasi-perceptual process, even simultaneously so. So it is just not a good objection to say that these examples yield modal beliefs but not non-modal ones.[78] Thirdly, I note that all three examples come directly from the literature and have been hotly debated, but the debate always concerned the *epistemic justification* of these beliefs rather than the question whether these beliefs are 'really' modal or non-modal beliefs. So, in any case, my analysis of these examples shall bear on this debate in the literature regardless of how these beliefs are most appropriately characterised.

(c) Finally, it may be argued that we should not call this a quasi-perceptual belief because the belief is not *sufficiently* determined by a quasi-perceptual experience, at least not as much as ordinary perceptual beliefs are determined by perceptual experiences: quasi-perceptual experiences are only directly involved in the *first* step of the two-step schema for

[78] Dorsch (2016b, pp.103–4) responds to a similar objection along similar lines.

quasi-perceptual beliefs (3.9) — they determine the propositional content — not in the *second* step — they do not determine the attitude of belief. This is true, and it is an important observation. But Dorsch (2016b, 102) responds:

> [T]he lack of a belief-like attitude does not deprive imaginative experiences of their potential to ground knowledge in an experiential way. Because the rational determination of the content and the rational determination of the attitude of beliefs are independent of each other, to an extent that they may even involve very different mental episodes or states, an experience without a belief-like attitude may still be central to the determination of the content of a given belief, even though it cannot play any role in the determination of its attitude.

I argue more thoroughly why this is the case in the next Section.

3.3.4 Why quasi-perceptual beliefs require meta-beliefs

I begin with an observation: *most* relevant propositions that come to mind on the basis of an ordinary perception are candidates for belief, while *very few*, if *any*, propositions that come to mind on the basis of a *quasi*-perception are candidates for belief. If you see an elephant in the room, then every aspect about this scenario can yield beliefs: e.g. that the elephant is happy or sad, that it is grey and hairless or brown and surprisingly hairy, that it is calm or shuffling around uncomfortably, etc. Which of these possibilities is the case depends on the world: it is there to be seen. But if you *imagine* an elephant in the room, then there generally are very few aspects of this scenario that than yield beliefs, even though many propositions may come to mind. You can imagine an elephant in the room as long as you want, but this will not yield the belief that there is an elephant in the room; you can imagine the elephant sad or happy, but this will not yield the belief that the elephant is sad or happy, and so on and so forth. You will only adopt the attitude of belief to propositions that come to mind on the basis of an imagined quasi-perception if you *believe*

that your imagined scenario accurately represents the world. While this will rarely be the case for the proposition that there is an elephant in the room, it might very well be the case for the proposition that your room is too small to comfortably fit three elephants in it.

The explanation for this difference between ordinary perception and imagined quasi-perception is evident: we *know* that the content of ordinary perception generally accurately represents the world (and we even know, if only instinctively, under which conditions it does so), and we *know* that the content of our imagination generally does *not* accurately represent the world. It is *imagination*, after all: imagination is 'epistemically innocent' and 'not world-sensitive', and we are generally well aware of it. If an act of imagination yields a belief about the natural world, then this belief must have been obtained at least partly in virtue of something *else* than imagination: in virtue of the meta-belief that the content of our imagination accurately represents the world.

The case of memory is somewhat of an intermediate case in between ordinary perception and imagined quasi-perception, and it nicely illustrates the need for meta-beliefs in quasi-perceptual belief-yielding processes. Some episodic memories may present themselves so convincingly that we do not hesitate for a moment to adopt the attitude of belief to the propositions that come to mind. But this is not always the case. Many propositions that come to mind on the basis of the episodic memory experience will (or *should*) make us pause and think: is my memory *accurate*? This difference can show itself even in the context of a single act of episodic memory: if you *remember* seeing an elephant, for example, you may have to pay a little more effort to remember the size of its ears than you have to remember the elephant's color — this will, of course, depend strongly on what you paid *attention* to during the creation of the memory.

To argue that processes of obtaining quasi-perceptual beliefs require meta-beliefs in the second step of (3.9), Dorsch (2016b, p.101–2) draws a helpful analogy with cases where we obtain beliefs through the perception of *realistic portraits*:

> When Henry VIII looked at Holbein's portrait of Anne of Cleves (see Figure [3.4]), he was able to acquire knowledge about her visual appearance, and not only recognitional [practical] knowledge enabling him to recognise her at their first meeting, but also propositional knowledge — such as the insight that she had brown eyes. That he endorsed the proposition *that Anne's eyes were brown* (and not that they were green, say) was determined solely by how his experience of the picture presented the eyes of the depicted woman as being. By contrast, that he came to *believe* this proposition (rather than merely entertaining or imagining it) was exclusively determined by his ancillary belief that Holbein's painting was an accurate portrait of Anne — a belief that could not be grounded in his experience of the picture (i.e. visual depictions do not tell us whether they are accurate). [...] So, when we form experience-based beliefs about the visual appearance of depicted people or objects, the content of our belief and the rational determination of its attitude are independent of each other and due to different factors: the first to our pictorial experience, and the second to our ancillary belief about the accuracy of the portrait (and the fact that it is a portrait in the first place).

We need not go back in history for illustrations of this phenomenon. In modern society, we gain most of our 'perceptual' knowledge by watching a phone, a computer screen or a TV. I can gain the knowledge that Asian elephants have smaller ears than African elephants by watching a Netflix documentary about elephants, because I *believe* that the documentary *accurately* represents elephants; but I will not gain this knowledge by watching a kid's cartoon show on elephants, because I believe that a kid's cartoon show cannot be trusted to represent elephants accurately. This example is entirely analogous to the case of Henry VIII gaining 'perceptual' knowledge about Anne of Cleves by observing her realistic portrait.

The same happens in the examples of novel beliefs gained through memory mentioned in Section 3.2.3. To make this more evident, consider the following example. Suppose that you have (what you believe are) two distinct episodic memories of seeing an elephant, but the elephants look rather different in each memory (perhaps you remember them from watch-

Figure 3.4: "Hans Holbein the Younger, *Anne of Cleves*, Detail (1539), The Louvre, Paris." Caption from (Dorsch, 2016b, p.101); image (colored version) courtesy of `wikipedia.com`.

ing the TV shows mentioned above); see Figure 3.5. Now, *which* memory will you be willing to use to form a novel belief about elephants? The answer is clear: only the memory of which you believe that it accurately represents real elephants. The same is the case for the three examples of novel beliefs gained through acts of imagination discussed in the previous Section: you can rationally determine novel quasi-perceptual beliefs on the basis of an act of imagination *only if* you have the meta-belief that your imagined scenario is an accurate representation of the world.

I note that this meta-belief is a belief, epistemically speaking, but it need not be an *occurrent* belief. The meta-belief may remain dispositional: in this case, it *will* become occurrent when its manifestation conditions are met (e.g. when you are asked a question about the accuracy of your quasi-perception) and it *does* influence behavior, at the very least in the sense that it motivates you to adopt the attitude of belief to a proposition (in step 2 of the two-step process for quasi-perceptual beliefs (3.9)). So, processes of rationally determining quasi-perceptual beliefs necessarily in-

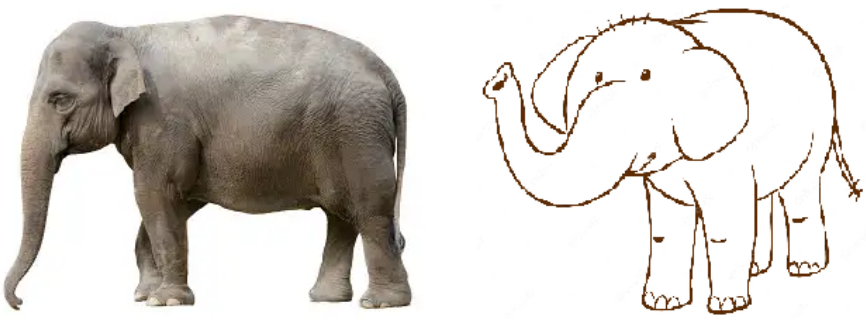

Figure 3.5: Accurate (left) and inaccurate (right) depiction of an elephant. (Images courtesy of freepik.com.)

volve meta-beliefs, whether occurrent or dispositional. This is important, epistemologically speaking, as I shall discuss in Section 3.5.

I flag that the meta-belief is necessary for *rational determination* of quasi-perceptual beliefs (3.9). This is the only 'epistemically valid' way of gaining quasi-perceptual beliefs that I consider. One may of course gain quasi-perceptual beliefs through means *other* than their rational determination (3.9), just like one may gain perceptual beliefs through means other than their rational determination (3.8). Desires, hopes and wild expectations, for example, may also be (partial) sources of perceptual belief — say, the perceptual belief that the person you fancy is looking at you, while they are actually not — but these beliefs are not *rationally determined* (and, hence, they can be regarded *unreliable* and *unrobust*). I therefore limit my attention to rationally determined quasi-perceptual beliefs, as only these beliefs have any chance of being reliable and robust, i.e. justified.

I next wish to note that step 2 of the two-step process for quasi-perceptual beliefs — the rational determination of the attitude of belief on the basis of a meta-belief — is, epistemologically speaking, an *inferential* process. It is an implicit or explicit inference from a meta-belief. If the meta-belief is occurrent, then the inference is explicit; if the meta-belief is dispositional, then the inference remains implicit. But, epistemologically

speaking, the inference is always there. This vindicates Stuart's remark (2021, p.1332), quoted in Section 3.2.1, that:

> If we are to do state-based epistemology of imagination, that is, if we are to find out how imagined mental states come to be known or play a role in gaining new knowledge, I suggest that the imagined content must figure somehow into an argument, inference, or other kind of *process*.

The idea that quasi-*perceptual* belief-yielding processes are partly *inferential* may sound objectionable to some: are quasi-perceptual belief-yielding processes not supposed to be *non*-inferential, like ordinary perceptual belief-yielding processes (3.8) are? To respond to this objection, I again give the word to Dorsch (2016b, p.107):

> While the content of the knowledge that we acquire in the examples is determined in a non-inferential way by the sensory (e.g. visual) content of our imaginative experience concerned, its attitude is determined inferentially on the basis of our ancillary [meta-]belief that we produced this experience in a reliable and safe (i.e. truth-preserving) way. This shows that the justification of experiential knowledge — that is, knowledge the content of which is the result of abstraction from, and conceptualisation of, sensory experience — can be partly inferential. If challenged, we should defend our [quasi-perceptual] belief [...] by reference to our imaginative visual experience [...] (the sensory element) and by reference to the reliability and safety of our imaginative project and the underlying perceptual or mnemonic experiences (the inferential element). In other words, experiential [quasi-perceptual] knowledge and inferentiality are compatible with each other — contrary to what is sometimes thought.

I flag that Dorsch (2016b) here writes that the meta-belief concerns the *reliability and safety*, i.e. robustness, of the quasi-perceptual belief-yielding *process*, which seems wrong: we not necessarily have a belief about the reliability of the quasi-perceptual *process* in the formation of quasi-perceptual beliefs. Dorsch is ambiguous about the content of the meta-belief; recall

that he also wrote (*ibid.*, p.102, quoted on 136) that: "the attitude of the resulting belief is, by contrast, determined by our [meta-]beliefs about the accuracy of how our imaginative experience visually presents things as being." This seems correct: we not necessarily have a belief about the reliability of the quasi-perceptual *process* in the formation of quasi-perceptual belief, but we *do* necessarily have a belief about the *accuracy of the result*.

In any case, quasi-perceptual beliefs are obtained in two steps (3.9): a quasi-perceptual 'experiential' step that determines the content of the to-be-believed proposition, and an inferential step that determines the attitude that we adopt towards the proposition: the attitude of belief. I repeat: despite the fact that quasi-perceptual beliefs are 'rationally determined' by a two-step process that is partly quasi-perceptual and partly inferential, quasi-perceptual beliefs are not *thereby* justified. Far from it, in fact. So let us now take a look at the justification of quasi-perceptual beliefs: what is *required* for it, and what is *responsible* for it.

3.4 Justifying quasi-perceptual beliefs

3.4.1 A criterion for justified quasi-perceptual beliefs

This bears repeating: the *rational determination* of quasi-perceptual beliefs (3.9) is not full-fledged *epistemic justification* (3.4) of them. You may have the meta-belief that your remembered or imagined scenario accurately represents the world, while it actually misrepresents the world: your meta-belief may be *false*, in which case the quasi-perceptual belief formed on the basis of this false meta-belief will be unreliable, i.e. *unjustified*. For memory, this is the case when we falsely believe that we are remembering while we are not actually remembering but rather 'merely imagining'. For imagination, it is the case when we falsely believe that our imagined scenario accurately represents the world.

I note next that, because the meta-belief required for the rational determination of quasi-perceptual beliefs concerns the relation between

a quasi-perceived scenario and the natural world, this meta-belief must be based on our background beliefs *about* the natural world. Given that the justification (3.4) of quasi-perceptual beliefs depends on the reliability and robustness of the belief-yielding process, *the justification of quasi-perceptual beliefs depends at least partly on the truth of these meta-beliefs.* We should recognise, then, that the justification of quasi-perceptual beliefs depends on the extent to which our background beliefs about the natural world are themselves *true beliefs*, perhaps even *knowledge*. This is what Gendler (1998, p.415) meant when she poetically wrote that:

> The justificatory force of thought experiments [and other epistemic acts of imagination] is thus parasitic on the extent to which the messy twisted web of background beliefs that underpin our navigation of the world are rightly considered knowledge.

This statement might give some readers pause. If the justification of quasi-perceptual beliefs gained through acts of imagination depends on, or "is parasitic on", on the truth and justification of our *background beliefs*, in virtue of *what*, precisely, are quasi-perceptual beliefs justified: in virtue of imagination, or in virtue of our background beliefs, or both? Both, as I will now argue in more detail.

Quasi-perceptual belief-yielding processes are *two-step* processes (3.9) that involve *both* quasi-perception (step 1) *and* implicit or explicit inferences on the basis of a meta-belief about the accuracy of our imagined scenario (step 2). Consequently, the justification (3.4) of quasi-perceptual beliefs resides in the reliability and robustness of *both steps*, not just in the reliability and robustness of either one step. So, to find the 'source' of justification of quasi-perceptual beliefs, we must find the 'source' of reliability and robustness in both step 1 and step 2 of the two-step process (3.9).

(For the sake of convenience, I focus my attention on reliability, and I lay aside considerations concerning *robustness*, as many, if not all, consid-

erations about reliability will apply equally well to robustness.)[79]

Fortunately, there is an interesting relation between the reliability of these two steps. Namely, the reliability of *both* step 1 and step 2 depend on the *same* two factors: (i) on the reliability of the quasi-perceptual process (step 1) and (ii) on the truth of the subject's relevant background beliefs about the natural world. I shall illustrate this with three examples.

Let us first look at two slight variations on Example 1 from Section 3.3.3, the fitting-elephants-in-the-room example. Suppose that you repeat this Example several times over the course of several days, to see whether your belief about the amount of elephants that fit in your room is reliable.

Suppose, first, that you are a highly skilled mental simulator but that you have false background beliefs about elephants: you can move around imagined elephants in your mind whilst carefully preserving their size and without creating overlap between the fitted elephants, and thus, every time you try to fit elephants in your room, the same proposition comes to mind — say, the *false* proposition that you can easily fit 5, but not 6, elephants in your room. Now, is this process reliable? This will depend on the truth of the meta-belief about the accuracy of your imaginings (step 2), and hence on the truth of your relevant background beliefs on which your meta-belief is based, i.e. on whether your imagined elephant has a size that *accurately represents* the size of real elephants. In this case, it turns out that your background belief about the size of elephants is false: your imagined elephants are much, much smaller than real elephants. Thus, your meta-belief that your imagined elephants accurately represent real elephants is false. In this case, then, your quasi-perception is unreliable (it yields false beliefs more often than not) *because* your background belief about the size of elephants is false. I take this scenario to be straightforwardly generalizable: if your background beliefs are false,

[79] The following seems true: if a belief-yielding process is robust and yields a *true belief*, then it is reliable. Beyond this, reliability and robustness seem independent: a belief-yielding process can be reliable and unrobust, as is the case when the process yields many different true beliefs; it can also be unreliable and robust, as is the case when the process often yields the same false belief.

then your quasi-perception will be unreliable despite you being a skilled mental modeller (step 1), and then your meta-belief is false. Thus, your quasi-perceptual belief is false, and the process of obtaining it is unreliable. False in, false out.

Let us now look at a scenario where the relation is the other way around: a scenario where your background beliefs are true but your quasi-perceptual process itself is unreliable. Suppose that you have true background beliefs about the size of elephants and about the size of your room, but that you are very bad at mental simulation, to the extent that your imagined elephants change size or overlap in various ways when you try to fit them in your room. Every time you try, a different proposition comes to mind upon trying to fit as much elephants in the room as you can: sometimes you can easily fit 4 elephants, but sometimes you can barely fit 3. In this case, your background beliefs were true, but your imagined manipulation of the quasi-perceived scenario was unreliable (and unrobust). Your imagined manipulation of the quasi-perceived scenario was not *truth-preserving*, as they say. Consequently, your meta-belief that your quasi-perceived scenario accurately represents the world is also false. Thus, your quasi-perceptual belief is false, and the process of obtaining it is unreliable (and unrobust). I also take this situation to be straightforwardly generalizable: if your background beliefs are true but your quasi-perception is unreliable because you are a faulty mental simulator, then your meta-belief will be false. Thus, your quasi-perceptual belief is false, and the process of obtaining it unreliable.

Now we return to the original version of Example 1. Suppose that you *do* have true background beliefs about the size of elephants and the size of your room, and that you *are* a skilled mental simulator, in the sense that your imagined fitting of elephants in your room is truth-preserving. In this case, your quasi-perceptual process is reliable (step 1), because your relevant background beliefs are true and the imagined manipulation of your quasi-perceived scenario is truth-preserving. The inference on the basis of your meta-belief about the accuracy of your imaginings is

reliable (step 2) because your meta-belief is true, which is the case because your relevant background beliefs were true and the imagined manipulation of your quasi-perceived scenario is truth-preserving. But these are the *same* criteria for reliability for both step 1 and step 2 of the two-step process for quasi-perceptual beliefs (3.9)! So: your quasi-perceptual belief-forming process (3.9) is justified iff your relevant background beliefs are true and the imagined manipulation of your quasi-perceived scenario is truth-preserving.

Are we done now? No: we should not forget the contribution of *episodic memory* to the justification of quasi-perceptual beliefs. Above, I discussed the justification of the quasi-perceptual belief gained in Example 1 from Section 3.3.3, which was an example that did not involve episodic memory. The content of many other acts of imagination, however, is heavily infused with mnemonic content. This was the case in Examples 2 and 3 from Section 3.3.3, In these cases, the justification of the quasi-perceptual belief also depends on the reliability of our relevant episodic memories, i.e. on our ability to truly episodically *remember* the scenario — in Example 2, our ability to remember, thus gain a true belief about, your friend hitting the drums a few times; in Example 3, our ability to remember, thus gain true beliefs about, the windows in each room of the house you grew up in. In these cases, our relevant episodic memories are *distinct* from our relevant background beliefs because the relevant features of our episodic memories *are not* propositional (background) beliefs: it is precisely the aim of the act of imagination in Examples 2 and 3 to *make* these features propositionally available *for* belief; recall my discussion of gaining novel beliefs through memory from Section 3.2.3. If our relevant episodic memories have already been 'exhaustively interpreted', in the sense that all relevant features of these episodic memories are *also* propositional background beliefs, then the condition that our episodic memories are reliable will collapse into the condition that our background beliefs are true. In general, however, this will not be the case, so we should keep these two conditions distinct.

There are thus *three* conditions for the justification of quasi-perceptual beliefs: (i) the relevant episodic memories must be reliable, (ii) the relevant background knowledge must be true, and (iii) the imagined manipulation of the quasi-perceived scenario must be truth-preserving. (Dorsch (2016b, p.96–99) arrives at the same conclusion, albeit phrased slightly differently.) Explication:

> [Justified Quasi-Perceptual Belief] Subject S's quasi-perceptual belief that p is *justified* iff (i) the episodic memories of S relevant for p are reliable, and (ii) the background beliefs of S relevant for p are true, and (iii) the imagined manipulation of the quasi-perceived scenario in step 1 of (3.9) is truth-preserving. (3.10)

With this criterion for the justification of quasi-perceptual beliefs (3.10) in hand, let us now return to the question posed at the top of this Section: in virtue of *what*, precisely, are quasi-perceptual beliefs justified? Is it in virtue of imagination or in virtue of our background beliefs, or both? From the discussion leading up to the criterion for justification of quasi-perceptual beliefs (3.10), the answer is evident: in virtue of both imagination (condition (iii)) and our background beliefs (condition (ii)), *and*, additionally, in virtue of the reliability of our relevant episodic memories (condition (i)). Thus it seems that quasi-perceptual beliefs are justified in virtue of the reliable and robust functioning of not one but *three* different cognitive faculties: imagination, reason and memory.

But things are more subtle than they appear, as will transpire next.

3.4.2 The Constraint Claim

I shall take it that the first two conditions in (3.10) require no further discussion. But condition (iii) *does* require further discussion: under which conditions is the imagined manipulation of a quasi-perceived scenario truth-preserving, or at least *reliable*, i.e. more often than not truth-preserving? Widespread consensus dictates that the imagined manipula-

tion of a quasi-perceived scenario *can* indeed be reliable, i.e. more often than not truth-preserving, under a certain condition. The condition is, perhaps unsurprisingly, that the act of imagination is *properly constrained to be 'reality-oriented'*, in the sense that (i) the content of the act of imagination (both the imagined set-up and imagined manipulations thereof) is constrained by (scientific) background *knowledge* of the natural world, and that (ii) the appropriate accepted modality for the act of imagination is *nomological* possibility,[80] thus guaranteeing that the content of our imagination accurately represents events that are possible in the natural world; see e.g. (Miščević, 1992, 2007, 2022; Nersessian, 1993, 2002, 2007, 2018; Schwartz and Black, 1999; Gendler, 2004; Nichols and Stich, 2000, 2003; Frigg, 2010b; Meynell, 2014; Dorsch, 2015; Langland-Hassan, 2016, 2020; Kind, 2016, 2018; Williamson, 2016; Salis, 2020; Myers, 2021; Williams, 2021; Hyde, 2021; Gauker, 2021; Badura and Kind, 2021; Stuart, 2021; Özgün and Schoonen, 2022; Berto, 2022, 2023; Miyazono and Tooming, 2023a,b).

At first sight, the notion of *imagining under proper constraint* is somewhat puzzling vis-à-vis the *voluntariness* inherent to imagination. Kind and Kung (2016, p.1) write:[81]

> How can the same mental activity that allows us to fly completely free of reality also teach us something about it? [This is] *the puzzle of imaginative use.*

But this puzzlement should be only temporary, as Kind (2016, p. 146) writes optimistically:

[80] Recall that in Chapter 2 I argued that *acceptance of a possibility in some appropriate modality* is a necessary condition for imagination (if the type of imagination is propositional (2.14), then the acceptance is occurrent; if the type of imagination is action-like (2.31), then the acceptance is dispositional.

[81] I wish to flag that Kind and Kung (2016)'s way of phrasing the *puzzle of imaginative use* — "how can the same activity that enables us to fly completely free of reality also teach us something about it?" — is not very poignant. At least it does not present well why this is an epistemological puzzle *unique to imagination*. Mathematics and logic, for example, also enable us to "fly completely free of reality" and they too can, and often do, "teach us something about it." Perhaps the puzzle of imaginative use is informatively analogous to the puzzle of the "unreasonable effectiveness of mathematics in the natural sciences" (Wigner, 1960). Someone (not me) should figure this out.

> The freedom we enjoy when imagining does not show that we must always proceed completely unfettered, and in fact it is our ability to constrain our imaginings in light of facts about the world that enables us to learn from them [...] [to the extent that] an imaginative project can play a *justificatory* role with respect to beliefs about the world.

I make this conditional claim explicit:

The Constraint Claim:
If the content of an act of imagination is properly constrained, then the act of imagination can yield justified quasi-perceptual beliefs. (3.11)

I note that the Constraint Claim specifies a *sufficient* condition for when quasi-perceptual belief-yielding processes yield *justified* beliefs. Most above-mentioned authors, except for rare exceptions such as Stuart (2020) and perhaps Gauker (2021), also consider it a *necessary* condition. (Stuart (2020) argues that unconstrained, "anarchic" imagination has also yielded scientific knowledge and understanding in the past.) For my purpose, it suffices to look at the Constraint Claim only as a sufficient condition.

The idea common to all above-mentioned adherents to the Constraint Claim (3.11) is that, even though in *theory* we can imagine whatever we want, in *practice* much of the content of our acts of imagination is *not* chosen freely. Quite the contrary: the content of our acts of imagination is often significantly *constrained* — by the (epistemic) purpose of our act of imagination; by what follows from the initial imagined premise; by what seems relevant and by what we deem possible; by our memories; by our habits, *aliefs*, emotions[82], primes, desires, expectations (recall Table 3.1); by our cognitive architecture; by what we *believe* and *know* about reality; and so on and so forth. In fact, constraints on imagination are numerous and come in a wide variety: they can be deliberate or indeliberate,

[82] Think of feeling resistance to imagine morally repugnant propositions; see 'the problem of imaginative resistance' (Gendler, 2000a; Gendler and Liao, 2016; Tuna, 2020), c.f. Kim et al. (2019).

voluntary or involuntary, conscious or unconscious, consistent or inconsistent, occurrent or dispositional; they can come alone or in combination, and they might reinforce or contradict each other. And surely *some* constraints or combinations of constraints, used at the right time, in the right way, with the right purpose, by the right person, are sufficiently *proper* to make sure that the content of our imagination *accurately represents the world*, thus enabling the justification of quasi-perceptual beliefs.

The Constraint Claim (3.11) poses an ostensible problem for the idea that *imagination* is a source of justification. While the notion of 'imagining under proper constraint' is appealing (as is evident from the large amount of adherents), the Constraint Claim poses a *prima facie* dilemma for proponents of the idea that it is *imagination* that crucially (or even substantially) contributes to the justification of quasi-perceptual beliefs. This dilemma was recently put forward and discussed by Kinberg and Levy (2022). I next discuss this dilemma and formulate my response.

3.4.3 A Dilemma for the Constraint Claim

The dilemma discussed by Kinberg and Levy (2022) aims at the distinction between *deliberate* and *indeliberate* constraints. Deliberate constraints on imagination are occurrently and voluntarily *chosen* constraints on the content of our acts of imagination; indeliberate constraints on imagination are constraints on the content of our acts of imagination that are *not* occurrently and voluntarily chosen.

To illustrate this distinction, let us again return to the Examples from Section 3.3.3, beginning with Example 1: fitting-elephants-in-the-room. Suppose that you have strong background beliefs about the exact size of elephants and the exact size of your room, and that you *carefully* use these background beliefs to determine the (relative) size of the imagined elephants and the (relative) size of your room. This is an example of a deliberate constraint on imagination. It may be argued that the term "deliberate constraint" is a bit misleading, as Kinberg and Levy (*ibid.*, p.10, fn.11) note rightfully that "the term 'constraints' appears too weak

for what several of the relevant authors have in mind. The suggestion seems to be that the imagination's content, and especially its output — in successful cases, that is — is not merely kept within certain bounds but is *positively determined* by the initial state given appropriate principles of change." While this is true, we should not forget that the content of imagination is *under-determined* by the choice of *topic*, i.e. it is not possible to choose and determine the *total* content of our imagination; recall Chapter 2, Section 2.4.2, Feature VIII. *Under-determination.* So, it is not a misnomer to call these deliberate and explicit choices that constrain, and to some extent positively determine, the content of our imagination *deliberate constraints.*

If, by contrast, you do *not* have strong background beliefs about the exact size of elephants and the exact size of your room but instead you are just 'winging it', and you imagine the (relative) sizes of elephants and your room as they 'come to you', then this is an example of an *indeliberate* constraint. Kinberg and Levy (2022) call such indeliberate constraints "black-box constraints": they are the constraints on our imagination of which we neither precisely know their source nor precisely understand the conditions for their reliable functioning (if there are any). Perhaps a clearer example of an indeliberate "black-box" constraint is Example 3, where you quasi-perceptually *counted* the amount of windows in the house you grew up in. The amount of windows in each house that you conjure up in your episodic memory is *not* occurrently and voluntarily chosen: every room just 'comes to you' in your episodic memory, allowing you to *count* the windows. If you knew the amount of windows in each room, then the content of your act of imagination would have been deliberately constrained — but then you would not have needed to count the windows anymore, because you already knew how many there are, which we assumed is not the case.

The dilemma for imagination as a source of justification, discussed by Kinberg and Levy (2022), then, runs as follows. Assume that the content of our acts of imagination is constrained either deliberately or in-

deliberately. If it is constrained deliberately, then the act of imagination may indeed yield justified quasi-perceptual beliefs, because the constraints make sure that the imagined manipulation of the quasi-perceived scenario is truth-preserving, which is condition (iii) in (3.10). But in this case, the beliefs are justified, not in virtue of imagination, but rather in virtue of the *constraints on* imagination, which are provided by something *else* than imagination, i.e. by *reason* or *memory*, or both, but *not* by imagination. If, by contrast, the content of our act of imagination is constrained indeliberately, i.e. if it is constrained by 'black-box constraints', then the act of imagination is necessarily unreliable and hence does not yield *justified* quasi-perceptual belief. In neither case, imagination is a source of justified quasi-perceptual beliefs.

I explicitly formulate this Constraint Dilemma as follows:[83]

The Constraint Dilemma:

Background Assumption: The content of an act of imagination is constrained either deliberately or indeliberately.

Horn (I): If it is constrained deliberately, then the act of imagination may yield justified quasi-perceptual beliefs, but then the beliefs are justified in virtue of the *constraints on* imagination, not in virtue of imagination itself. (3.12)

Horn (II): If it is constrained indeliberately, then the act of imagination does not yield justified quasi-perceptual beliefs.

Dilemma: In neither (I) nor (II), imagination is a source of justified quasi-perceptual beliefs.

I am unconvinced by the Constraint Dilemma (3.12) and I shall argue

[83] Kinberg and Levy (2022) discuss their Dilemma with respect to *knowledge* as a whole, not with respect to only justification. However, it is clear that the target of Kinberg and Levy's Dilemma is justification, as they themselves acknowledge (p. 4) that "the issue at hand is not whether the imagination is *in some way or other involved* in knowledge acquisition. Rather, the question concerns the warrant-providing [i.e. justifying] role of the imagination." I thus feel licensed to limit the scope of Kinberg and Levy's Dilemma to justification.

against it in the next Sections. Here I wish to note what the Dilemma *does* achieve: the Constraint Dilemma shows us that we cannot straightforwardly conclude from the criterion for justification of quasi-perceptual beliefs (3.10) that, due to condition (iii) in (3.10), quasi-perceptual beliefs are justified at least partly in virtue of *imagination*. The issue is more nuanced: if imagined manipulation of a quasi-perceived scenario is truth-preserving, then we must determine very carefully *in virtue of what* it is truth-preserving — in virtue of imagination itself or in virtue of the (non-imaginative) *constraints on* imagination.

I next put forward a three-fold response to the Constraint Dilemma (3.12). I argue that: (i) Horn I of the Dilemma is *false* (Section 3.4.4); (ii) Horn II of the Dilemma is *false* (Section 3.4.5); and (iii) the Constraint Dilemma is a *false* dilemma (Section 3.4.6). I conclude, *contra* Kinberg and Levy (2022), that quasi-perceptual beliefs can be justified at least partly *in virtue of imagination*. I then relate my response to the Constraint Dilemma to the arguments put forward by Miyazono and Tooming (2023a), who respond to a dilemma that is similar (but not identical) to the Constraint Dilemma along similar (but not identical) lines as I do.

3.4.4 Against Horn I: Thought Experiments

Horn I of the Constraint Dilemma (3.12) states that, if the content of our act of imagination is constrained deliberately, then it may yield justified quasi-perceptual beliefs, but then the beliefs are justified in virtue of the *constraints on* imagination, not in virtue of imagination itself. Kinberg and Levy (2022, pp. 11–12, our emphasis) elaborate:

> For under this construal, the imagination appears to serve as no more than an arena, as it were, for performing "regular" hypothetical inferences. What it does is, essentially, to put forward a proposition and explore whether it leads to some consequence of interest. And, crucially, *it is the quality of these inferences, and not the imaginary setup, that justifies us in believing their output.* To be sure,

> these inferences are performed in the imagination — but that appears unimportant, epistemically speaking. Just as we do not want to speak of "knowledge via paper" when we perform a calculation with pencil and paper, or of "knowledge via blackboard" when our instruments consist of a chalk and a blackboard (or, for that matter, of "knowledge via Mac" if that is the machine we're using), so *it does not seem appropriate to speak here of a special form of knowledge via the imagination.*

Thus, Kinberg and Levy claim that, *regardless* of the details of the *psychological* (or even physical) process of obtaining justified quasi-perceptual beliefs (the beliefs may be obtained via imagination, or via a drawing, or via computer simulation, etc.), *if a belief-yielding process is reducible to reasoning through an argument, then the beliefs yielded through this process are justified, epistemically speaking, only in virtue of the quality of the inferences that it would take to reach the conclusion via argument*, not in virtue of the details of the psychological process that actually unfolded and yielded the quasi-perceptual belief.

To begin, I wish to flag that this claim is neither new nor popular. In the literature on thought experiments — the paradigmatic example of deliberately constrained acts of episodic imagination — a similar claim, under the guise of the *the argument view* of thought experiments, has long been defended by Norton (1993, 1996, 2004a,b). Norton argued that thought experiments are, epistemically speaking, *nothing but arguments*, in exactly the sense that "their success is attributable [only] to the proper application of appropriate principles of inference" (Kinberg and Levy, 2022, p.12). Norton (2004b, p.44–52) writes:

> In so far as they tell us about the world, thought experiments draw on what we already know about it, either explicitly or tacitly. They then transform that knowledge by disguised argumentation. [...] If thought experiments can be used reliably epistemically, then they must be arguments (construed very broadly) that justify their outcomes or are constructible as such arguments.

Barely anyone in the contemporary literature is convinced by Norton's reductive *argument view*, however, specifically after it was thoroughly scrutinised and several distinct claims were identified in his position, some of which are much less plausible than others; see notably Brendel (2018) and Chapter 4, Section 4.2.5 of this Thesis. For the present purpose, I highlight two central claims (Brendel, 2018, p. 283), both of which appear to be endorsed by Kinberg and Levy in the quote at the top of this Section:

> **Norton's Reliability Thesis:** "If thought experiments can be used reliably epistemically, then they must be arguments (construed very broadly) that justify their outcomes or are constructible as such arguments" (Norton, 2004b, p.52). A thought experiment is a "reliable mode of inquiry" only if the argument into which it can be reconstructed justifies its conclusion.
>
> **Norton's Epistemic Thesis.** Thought experiments [and other deliberately constrained acts of episodic imagination] and the arguments associated with them have the same epistemic reach and epistemic significance. Thought experiments are not [ever] epistemically superior to their corresponding non-thought-experimental arguments — and vice versa. [...] a thought experiment epistemically justifies its outcome to the same degree as its associated argument justifies its conclusion.

The first problem with Norton's Theses is that Norton's conception of 'argument' is construed *too* broadly because, upon being pressed, Norton's conception of 'argument', which was limited to deductive arguments at first, quickly grew so wide that it now includes implicit inductive inferences, abduction, metaphorical reasoning, and even reasoning on the basis of informal logics and what have you. In response, Meynell (2014, 4154–5) writes:

> [B]roadening the conceptual category "argument" to include any source of knowledge beyond immediate perception of the object of

> knowledge risks deflating the category itself as a useful classification. For Norton's project to be useful and plausible he needs to show that TEs are arguments, not just that they can be reconstructed into arguments, while at same time construing "argument" sufficiently narrowly so as not to make the position trivially true. The claim that TEs are arguments must mean something more substantive than that they are linguistic objects with the function of persuading the hearer of something."

Amen.

The second, bigger problem for Norton's Theses is that they are demonstrably false. After identifying Norton's Theses, Brendel (2018, p. 289) describes how, notably, proponents of the so-called *mental modeling view* of thought experiments (references in Chapter 4, Section 4.2.5) argue that "the sort of reasoning that is operative in belief-formation processes based on contemplating imaginary scenarios is cognitively and epistemically different to the sort of reasoning that is going on when we infer a conclusion from premises in an argument", and that "there are some cases where 'the imagery' is epistemically crucial", as (Gendler, 2004, p. 1161) puts it, [because] imaginary scenarios evoke "quasi-sensory intuitions that could lead us to form new beliefs via a 'quasi-observational' imagistic kind of reasoning". In other words, these authors argue that Norton's Reliability Thesis is false:[84] thought experiments are experiments performed in the imagination, rather than in the laboratory, and they are not (always) epistemically reducible to pieces of reasoning.

How can mental imagery be epistemically crucial in a way that propositional content (the content of our mental states when we reason explicitly through arguments) cannot? The answer resides in the property that mental imagery has but propositional content does not: mental imagery can be epistemically crucial in a way that propositional content cannot if the (functional analog of) *spatial properties* of mental imagery play a crucial role in the "imagistic kind of reasoning" and, consequently, in the forma-

[84] In Section 3.4.7, I shall discuss how these authors also argue that Norton's Epistemic Thesis is false.

tion of a quasi-perceptual belief. It is in *this* sense that the inferences that we perform in the imagination can be epistemically distinct from inferences that we perform while reasoning — it is in *this* sense that it *does* "seem appropriate to speak here of a special form of knowledge via the imagination" (Kinberg and Levy, 2022, p.12), contra to what Kinberg and Levy claim (recall the quote at the beginning of this Section).

To see this more clearly, consider Gendler's (2004, p.1158) discussion of the fitting-elephants-in-the-room example that I introduced in Section 3.3.3, which was roughly copied from Gendler (2004):

> Were the beliefs you formed on the basis of your reasoning in each of these cases formed as the result of inference from known premises to inductively or deductively implied conclusions? A "yes" answer is most plausible in the case of our four elephants. Arguably, even before engaging in the reasoning process described, you had the justified true belief that elephants are of thus-and-such size, the justified true belief that the living room is of thus-and-such size, a set of justified true beliefs concerning the solidity and limited malleability of elephants and living-room walls, a set of justified true beliefs concerning the possible configuration of objects in spaces governed by Euclidian geometry, and so on. On the basis of these (perhaps tacit) beliefs, you engaged (again, perhaps tacitly) in a process of deductive reasoning which led you to the realization that four elephants would not, as a matter of fact, fit comfortably into your neighbor's living room. [...] But is that really what happened? My inclination is to think not. Rather, what happened is that you formed a judgment on the basis of your manipulation of your mental image, and — using that *new [spatial] information* — went on to draw your conclusion about the more general statement for which you took it to be evidence.
>
> If you are still unpersuaded, think about the following cases. Suppose that I had, instead, given you a piece of graph paper and a pencil, and asked you the same question, which you answered on the basis of a sketch that you made: would that be a case where you engaged in a process of deductive reasoning from known premises to a novel conclusion? Or suppose I had given you a three-dimensional

> scale-model of the room, along with four similarly scaled plastic elephants (and suppose it wasn't immediately clear whether or not the elephants could be placed comfortably therein): wouldn't you proceed by putting the elephants into the room, and *seeing* whether they fit? Suppose I took away the third and fourth elephants before you managed to place them in the room. Would your imaginary continuation of the process you had begun really be a process of *deductive reasoning*? [...]
>
> Wasn't it instead as if you performed an *experiment-in-thought*, on the basis of which you got some new information about your own judgments, which (perhaps because of tacit beliefs that you hold) you took to be relevant data in answering the question at hand?

So: Gendler argues that it is not the case that, if the content of our imagination is deliberately constrained, then the process of obtaining a quasi-perceptual belief reduces entirely — that is, both psychologically and epistemologically — to 'reasoning through an argument'. Let us briefly look at one oft-cited example that Gendler (*ibid.*, pp.1159–60) uses to support her case:

> Research by Roger Shepard and others has shown that judgments about topological similarity are generally made after engaging in the mental manipulation of an image: the greater the degree of rotation required to project one onto the other, the longer it takes to judge whether two figures are isomorphic (Shepard and Metzler, 1971; Shepard and Cooper, 1986). Here, as above, it seems that the reasoning process is quasi-perceptual: I observe something, and on the basis of my observation conclude something. While this latter step may be construed as inductive reasoning, it is hard to see how the former step could be construed as either inductive or deductive. It's true that the geometrical constraints which my reasoning process tracks *deductively imply* the conclusion I draw — but that doesn't mean that what I did was to reason deductively from known premises.

So, to find out why Horn I of the Constraint Dilemma is false, all we need to do is look at the psychological description of a quasi-perceptual

belief-yielding process and ask ourselves three questions: (i) is this psychological process wholly inferential, yes or no?; (ii) if no to (i), is the process crucially quasi-perceptual, in the sense that spatial properties of mental imagery play a crucial psychological and epistemic role? (iii) if yes to (ii), is the obtained belief nonetheless reliable? If yes to (iii), then the justification of the belief is at least partly due to imagination. The examples above demonstrate that this can be the case. Thus Horn I of the Constraint Dilemma is false. Given that Kinberg and Levy themselves also adopt a criterion for justification similar to my reliability-and-robustness criterion for justification (3.4),[85] they are also forced to accept that Horn I of the Constraint dilemma is false.

I shall expand on this line of thought in the next Sections, particularly Section 3.4.7. I here merely wanted to show that, with Horn I of their Dilemma, Kinberg and Levy (2022) appear to put forward Norton's Reliability and Epistemic Theses, which have already been scrutinised and contested in the literature, and currently have very few supporters. Kinberg and Levy appear to be unaware of this work concerning scientific thought experiments and the consensus reached: by putting forward essentially Norton's Reliability and Epistemic Thesis as Horn I of their Constraint Dilemma, they have taken a wrong turn.

3.4.5 Against Horn II: Experts in Imagination

Horn II of the Constraint Dilemma states that, if the content of our episodic imagination is constrained indeliberately, then it is necessarily unreliable and hence does not yield justified quasi-perceptual beliefs. I noted that Kinberg and Levy (2022) call such indeliberate constraints 'black-box constraints', which, they argued, result in (near) *necessarily*

[85] Kinberg and Levy (2022) write: "We do, in general, assume that reliability matters for knowledge, at least insofar as lack of reliability can defeat knowledge. We take this to be a modest and very plausible assumption, unlikely to be contested by anyone in this debate." Strictly speaking, this only holds that reliability is *necessary* for justification, not that it is *sufficient*. From their discussion throughout the paper, however, it is evident that Kinberg and Levy (2022) also consider reliability a *sufficient* condition for justification.

unreliable acts of imagination because we know and understand neither their source nor the conditions for their reliable functioning (if there are any).

Kinberg and Levy (2022) focus their discussion of Horn II on two often-discussed cases of non-deliberately constrained episodic imagination: (a) everyday exercises of the imagination, mostly pertaining to the prediction of simple physical events involving simple spatio-temporal matters ('Does this couch fit through that door?', or 'Will this tower of block fall over?'), i.e. so-called *intuitive physics*; and (b) evolutionary arguments in favor of the truth-conducive character of the cognitive mechanisms responsible for imagined manipulation of quasi-perceptual scenarios. I shall respond to both.

The content of our imagination can be constrained non-deliberately — i.e. involuntary or unintentionally — by a great number of factors: by our current moods, desires and emotions; by our immediate environment; by our cognitive habits and physiological architecture, i.e. by our "perceptual and motor systems" (Nersessian, 2018, p.317); by what we instinctively deem relevant, possible and conceivable; by our 'conceptual apparatus' (Kuhn, 1977); by our (implicit or explicit) background knowledge, beliefs and 'aliefs' (Gendler, 2008), by our memories; and what have you. Some of these constraints, most notably our current moods, desires and emotions, are evidently not truth-conducive. These non-deliberate constraints do not enable imagination to function as a reliable source of knowledge.

Other constraints, however, *are* often argued to be truth-conducive, notably our perceptual and motor systems, our background knowledge, and, of course, our memories. Kinberg and Levy (2022) focus their discussion on the truth-conduciveness of our perceptual and motor systems, so I shall discuss these first.

Many proponents of the idea that imagination is a source of justified beliefs buttress their claim by appealing to quotidian examples of every-day and down-to-earth uses of imagination of immediate practical relevance: will this couch fit through that doorway? (Dorsch, 2016b); will

this poem fit on the tune of that song? (*ibid.*); is this tower of bricks likely to fall over? (Myers, 2021); can I successfully jump over this river? (Williamson, 2016); how should I climb this cliff when wishing to avoid falling to my death? (*ibid.*); will I enjoy living in this house? (*ibid.*). In trying to find answers to these questions by deploying our imagination, we make indeliberate yet heavy-duty use of our *perceptual and motor systems* in constructing and manipulating realistic scenarios in our imagination. And, crucially, we often seem to be *successful* in obtaining the right answers. This led Williamson (2016, p.113, my italics) to argue that:

> *Far from being the opposite of knowing, imagining has the basic function of providing a means to knowledge* — and not primarily to knowledge of the deep, elusive sort that we may hope to gain from great works of fiction, but knowledge of far more mundane, widespread matters of immediate practical relevance.

There is extensive empirical research into the psychological underpinnings of mental simulation that crucially employ our perceptual and motor systems; see e.g. references in (Nersessian, 2018, §2); (Gendler, 2004, §4), (Miyazono and Tooming, 2023a). Kinberg and Levy focus their discussion on so-called *intuitive physics*, which is our "untutored ability to predict simple physical events" (Kinberg and Levy, 2022, p.5); c.f. Kubricht et al. (2017). Kinberg and Levy point to the fact that 'untutored' humans are generally *quite bad* — that is, unreliable and unrobust — at predicting the unfolding of simple physical events, because *our perceptual and motor mechanisms responsible for making these predictions are essentially a hodgepodge of evolved heuristic biases rather than genuinely truth-conducive mechanisms*: we are error-prone 'noisy Newtonians', as they say. Indeed, in general, it is well-known that the evolved *fitness* of our cognitive faculties does not imply that they are *truth-conducive*, see Sage (2004); Prakash et al. (2021); but c.f. Boudry and Vlerick (2014).

It seems that we now have a paradox in our hands. On the one hand, we seem to use our imagination constantly and successfully for predicting the future and, in many down-to-earth yet epistemically interesting ways,

and we thereby gain knowledge — thus thereby gain justified beliefs. On the other hand, there is extensive empirical evidence that the mechanisms that enable us to make these predictions in the imagination are generally quite unreliable and, hence, do not yield justified beliefs. What gives?

To begin, I note that Kinberg and Levy (2022) do not argue that indeliberately constrained imagination in the context of intuitive physics is *necessarily* unreliable. But they *do* argue that it is *overwhelmingly* unreliable, while additionally noting the following (2022, p.6):

> Now, perhaps a domain or a set of contexts wherein our imaginations are trustworthy can be systematically circumscribed. To the best of our knowledge, such a domain has yet to be delineated. And even if such a delineation were to be carried out, it is doubtful that it would encompass anything like the imagination *simpliciter* or that it could be generalized in any substantial way.

I shall argue next that there *is* a "domain or a set of contexts wherein our imagininations are truthworthy can be systematically circumscribed".

Undeniably our *untutored* ability to predict physical events is unreliable, however simple these events may be. But Kinberg and Levy (2022) — and many others — are mistaken to look for reliable uses of imagination in the context of 'intuitive physics', that is, in our *untutored* ability to predict physical events. What we should look for when we search for reliable uses of imagination are, of course, *tutored* uses of the imagination; that is, uses of imagination by those who have plenty of *experience* with trying, *successfully and unsuccessfully*, to gain knowledge in similar scenarios, whose (implicit or explicit) *memories* of these experiences are themselves reliable and whose relevant background beliefs are true — recall the criterion for the justification of quasi-perceptual beliefs (3.10). To return to an example mentioned in passing above, suppose you are about to climb a dangerous cliff and are trying to figure out — using your imagination — how you should reach the top without falling to your death. In which case should you trust your *own* judgment more: (1) when you have never climbed a cliff before, or (2) when you are a professional rock-climber? The

correct answer is obvious. To use another example (that does not involve relevant ordinary perceptions during the act of imagination): suppose that you have only ever seen an elephant once, in a zoo, from a distance, but that you live together with the zookeeper that takes care of the elephants at your local zoo. Whose judgment would you trust about the amount of elephants that would fit in your room: your own judgment, or your friend the zookeeper's? Here, too, the answer is obvious.

Mutatis mutandis for predicting physical events about the behavior of solid 'Newtonian' objects. Persons whose job or interest does not involve working with or moving solid bodies will be poor generators of knowledge about such events, whereas persons whose job or interest involves working with or moving bodies, ranging from mechanics, lorry drivers and practitioners of sports with objects (usually balls) to teachers of courses on classical mechanics, will be far better in generating knowledge — that is, far better in *reliably* yielding true quasi-perceptual beliefs — about the motion of solid objects in specific situations, especially when these events fall within their scope of experience and knowledge. Their imagination yields knowledge in situations where they are *knowledgeable*: they can *see* it, just like that, without going explicitly through chains of reasoning.

Let me illustrate this last point with an example that is relatively famous on the Internet, but which I have not encountered in the literature. Suppose that you have a metal disk with a hole in the middle, and you proceed to heat up the metal disk uniformly; see Figure 3.6. Because the metal disk heats up, the disk will expand (uniformly) due to thermal expansion. Now the question is: what will happen to the radius of the *inner* hole? Will it become larger or smaller, or will it stay constant?

As testified by responses scattered on the Internet, which I anecdotally confirmed by asking this question to my colleagues in my philosophy department, many will answer this question wrongly and say that the inner radius of the metal disk becomes *smaller*. Presumably, the fact that this question is often answered wrongly as such is that most of our day-to-day experience with expanding objects concerns *soft* objects rather than *solid*

Figure 3.6: Uniformly heating a holey metal disk.

objects. When you bake a donut, for example, the inner hole of the donut will become smaller as it expands in the baking process. But this is not the case for solid objects such as a metal disk. The correct answer is that the inner radius of the metal disk becomes *bigger*, not smaller. (Due to thermal expansion, *all* molecules of the metal disk increase their relative distance to each other, so too the ones on the edge of the inner hole: the inner ring must expand.) Interestingly, most *car mechanics* can easily answer this question correctly, because they know that, if the braking disk of a car is stuck on the driveshaft, then they should heat it up, so it will come off much more easily. Their knowledgeability of analogous cases can be directly used in this act of imagination to *indeliberately* (because non-occurrent or involuntary) constrain the content of their imagination and yet *reliably* yield a true quasi-perceptual belief.

All things considered, I claim that non-deliberately constrained imagination *can* be reliable in some domain, and that the reliability of non-deliberately constrained imagination depends on *relevant experiences*, i.e. those about situations comparable to the newly presented ones, and the reliability of our *memories* and the truth of our *background beliefs* of them.

For the sake of clarity, I make this claim explicit:

Reliability Claim.
Given some quasi-perceived scenario, *if* subject S has (i) ample experience with gaining knowledge in comparable scenarios, and (ii) S's relevant episodic memories are reliable, and (iii) S's relevant background beliefs are true, *then* it is possible that S performs an indeliberately constrained act of imagination that yields reliably true, i.e. justified, quasi-perceptual beliefs. (3.13)

Like the Constraint Claim (3.11), the Reliability Claim is only a claim about three *sufficient* conditions for indeliberately constrained yet reliable imaginings. I do not argue here that these three conditions are also jointly *necessary* for indeliberately constrained yet reliable imagining. Perhaps there are other jointly-sufficient conditions, which are beyond my present scope.

I take it that, if (i), (ii) and (iii) hold for an imagining subject in some context, then the subject is an *expert* in that context (recall the examples above). In other words, then, the Reliability Claim (3.13) states that 'distinctively imaginative justification' of quasi-perceptual beliefs obtained via indeliberately constrained acts of imagination is possible, while limiting its domain to *experts* in whatever relevant context. This provides a stark contrast between justified *quasi*-perceptual beliefs and justified *ordinary* perceptual beliefs, as ordinary *perceivers* have domains accessible to non-experts (e.g. perceptual beliefs of simple demonstratives such as There is a cat on the mat), as well as domains accessible to experts (complex, non-inferential perceptual beliefs), while *imaginers* only have epistemic domains accessible to experts. Although the Reliability Claim (3.13) seems often implicitly acknowledged, I am not aware of any explicit formulations of it in recent literature on the epistemology of imagination. This is especially surprising because Kuhn (1977, p.265) already expressed a similar sentiment long ago, albeit limited to the context of

scientific thought experiments (and given the assumption that scientists are *experts* in their respective discipline):

> [In a thought experiment,] the imagined situation must allow the scientist to employ his usual concepts in the way he has employed them before. It must not, that is, strain normal usage. [...] [Moreover, the] conflict that confronts the scientist in the [imagined] experimental situation must be one that, however unclearly seen, has confronted him before. Unless he has already that much experience, he is not yet prepared to learn from thought experiments alone.

I next make five systematic remarks about the Reliability Claim (3.13).

First, given that the way our perceptual and motor systems govern real-world action and constrain the content of our imagination crucially depends on the sum-total of our *experience* with using these systems in daily life, an important question pertaining to the Reliability Claim (3.13) is to what extent our background beliefs, memories and our 'perceptual and motor systems' are conceptually and functionally *distinct*. I cannot and shall not dive deeply into this thorny issue, but I note that proponents of the view that our perceptual and motor systems can constrain our imagination in a truth-conducive manner often attribute, at least partly, the justificatory force of these constraints to memory. For example, Nersessian (2018, p.310, my italics) writes that

> Thought experimenting is a species of reasoning rooted in the ability to imagine, anticipate, visualize, and *re-experience from memory*.

See also Mach (1897); Gendler (1998); Kuhn (1977).

Second, Kinberg and Levy might want to point out that the Reliability Claim (3.13) *shows* that knowledge is never 'distinctively imaginative', precisely because the reliability of our imagination — hence its justificatory force — should be attributed to the reliability of our *memories*, not to any aspect inherent to imagination. While this is true enough, I respond that imagination can *make use of memories in distinctively imaginative (and non-inferential) ways*, and it is precisely the *freedom* of imagination

that enables us to imagine scenarios that enable us to pay *attention* to — hence, *learn from* — features of our episodic memories in ways that we did not — and, perhaps, could not — pay attention to before. I illustrated this in particular with Examples 2 and 3 from Section 3.3.3.

Third, expertise is always relative to a certain context. Within this context, expertise secures the reliability of beliefs obtained in this context. But *outside* these contexts, in contexts that seem to us as similar but which are, in fact, significantly distinct, expertise can also be *deceptive* and may lead us astray. The above-mentioned example of the expanding Holey Metal Disk illustrates this well: if someone is an 'expert' with respect to expanding *soft* objects, then they might wrongly attempt to apply this expertise in the case of the Holey Metal Disk, which is *not* a soft object. This explains to a large extent how imagination can be *fallible*, how we can *trick ourselves* into falsely believing that our imagined scenario accurately represents the natural world — specifically, a part of the world about which we are experts — while it actually does not. This explains how we *rationally* determine *false* quasi-perceptual beliefs.

Fourth, all authors on imagination underwrite the connection between imagination and possibility. Recall of characterizations of imagination like: "imagination represents ways the world might be" (Huemer, 2001, p.54), or "to imagine something is to think of it as possible" (White, 1990). The Reliability Claim (3.13) seamlessly fits this direct connection between imagination and possibility. Whether imagination is a reliable source of true beliefs about the world depends on how remote the possibility we imagine is. The context of imagination matters a lot. When the imagination is running on steroids about possible worlds where the laws of nature are suspended and magic reigns, e.g. in fairy tales and other fantasies, mental states of imagination will rarely be reliable sources of true beliefs about the world. When the imagination is about quotidian possibilities, however, like imagining whether a couch will pass the doorway, or whether I shall reach the other side of the ditch when jumping, imagination will likely be a reliable source of true beliefs. More remote

possibilities are less constrained, whereas nearby possibilities are heavily constrained. This also ties in with expertise: imagining remote situations will amount to imagining situations incomparable with what the imaginer has ever experienced, and hence expertise by previous experience is non-existent.

Fifth, note that the three conditions in the antecedent of the Reliability Claim (3.13) differ in only one respect from the three conditions in the criterion for the justification of quasi-perceptual beliefs (3.10). Thus, substituting the Reliability Claim (3.13) into the criterion for justified quasi-perceptual beliefs (3.10), we obtain the following sufficient conditions for the *justification* of beliefs obtained through indeliberately constrained quasi-perceptual belief-yielding processes: subject S's quasi-perceptual belief that p, obtained via an *indeliberately constrained* act of imagination, is *justified* if (i) the episodic memories of S relevant for p are reliable, and (ii) the background beliefs of S relevant for p are true, and (iii) S has ample experience with gaining knowledge, and failing to gain knowledge, in relevant scenarios. The condition for justification (3.4) that the imagined manipulation of the quasi-perceived scenario is truth-preserving, or at least *reliable*, is thus secured by the condition that S has ample experience with gaining knowledge, and failing to gain knowledge, in relevant scenarios. This guarantees that S *knows how* to successfully gain knowledge, and which epistemic pitfalls to avoid, in the relevant type of scenarios, thus guaranteeing that their act of imagination is reliable.

3.4.6 Against the False Dilemma

Dilemmas are exclusive disjunctions between two horns. I claim that what Kinberg and Levy are serving as an exclusive disjunction, between deliberate and non-deliberate acts of imagination, is, in fact, an inclusive disjunction, and therefore they have served a false Dilemma.

Specifically, I wish to insist that *all acts of imagination are indeliberately constrained to some extent*, even if they are *also* deliberately constrained. There is no such thing as exclusively deliberately constrained

imagination: the content of our imagination is either constrained entirely indeliberately (when we 'attempt' to let our imagination 'run free' entirely, the content of our imagination is still indeliberately constrained by our memories, background beliefs, primes, moods, emotions, direct environment, etc.) or it is constrained both deliberately *and* indeliberately.

This, again, is a direct consequence of the *under-determination of imagined content*: a deliberate choice of *topic* for our imagination underdetermines the *content* that we actually imagine; c.f. (Langland-Hassan, 2016; Berto, 2022). Suppose, for example, that you *deliberately* choose to imagine, as realistically as possible, seeing a gorilla in the staircase. This is an instance of deliberately constrained imagination. But *how* you imagine this gorilla will depend on non-deliberate constraints mentioned in the previous Section, some of which are truth-conducive and some of which are not. You deliberately imagine *how* the gorilla is present in the staircase: sitting, standing, climbing, walking, lying, resting, moving his arms, looking angry or at peace, etc. You will imagine a black or grey animal, not a pink or purple one. You will imagine a gorilla that resembles gorillas from memory. After all, you are supposed to imagine a *gorilla*, and not a specimen of any other animal species. This interplay of deliberate and non-deliberate features is endemic in every act of imagination, or so I like to imagine. Expressing this same sentiment, Stuart (2021) urges us to explore the possible *combinations* of deliberate and non-deliberate constraints that might be epistemically relevant. I very much endorse this suggestion. The Constraint Dilemma is a false dilemma.

To recapitulate: indeliberately constrained imagination can be a source of justified quasi-perceptual beliefs, but only in so far as the imaginer is an expert about what is imagined, and only in so far her memory is reliable and her (implicit) background beliefs are true. This is the Reliability Claim (3.13). In nearly every act of imagination, deliberate and non-deliberate constraints figure, which makes the Constraint Dilemma (3.12) between them, as served by Kinberg and Levy (2022), a false dilemma. Horn I of the Dilemma turned out to be Norton's Reliability and Epistemic

Theses, which have been rejected by philosophers of thought experiments, which has apparently passed by Kinberg and Levy. Horn II of the Constraint Dilemma is false when stated generally, and becomes true when conditioned: if imagination is constrained by experts with reliable memories, then imagination can function as a source of justified beliefs.

3.4.7 Another dilemma: proper constraint and otherwise-inaccessible constraints

I next briefly discuss a dilemma put forward by Miyazono and Tooming (2023a), which is similar (but not identical) to the Constraint Dilemma (3.12) from Kinberg and Levy (2022), and they respond to this dilemma in a way that is similar (but not identical) to my response to the Constraint Dilemma.

Unlike the Constraint Dilemma (3.12), the dilemma discussed by Miyazono and Tooming (2023a) does not aim at the distinction between deliberate and indeliberate constraints, but rather aims at the distinction between *proper* and *improper constraint*. Recall (Section 3.4.2) that imagination is *properly constrained* if it is constrained (deliberately or indeliberately) in ways such that the content of our imagination represents the natural world. Miyazono and Tooming (2023a) elaborate:

> [I]f imagination can epistemically contribute to some justification or knowledge, it is in virtue of the fact that imagination is "properly constrained" in the sense that it is sensitive to the real features of the world (or, is "closely guided by reality as it is" (Kind, 2016, p.150), "governed by the world" (Kind, 2018, p.243), "constrained by the sort of things one would expect to see" (Langland-Hassan, 2016, p.70), "in some sense subject to constraints that in some sense reflect the state and structure of the world" (Williams, 2021, p.69).

In my own terms, imagination is properly constrained iff it is constrained (deliberately or indeliberately) by our background beliefs and memories, and the appropriate accepted modality in the act of imagination is *nomological* possibility, and thus the content of our imagination should represent

events that are *possible* in the real world.

The dilemma discussed by Miyazono and Tooming (2023a) now runs as follows. An act of imagination is either properly constrained or it is not properly constrained. If an act of imagination is properly constrained to the extent that it may yield *justified* quasi-perceptual beliefs, then the act of imagination does not *generate* novel justification for quasi-perceptual beliefs but instead *preserves* prior justification provided by the (non-imaginative) constraints. If an act of imagination is *not* properly constrained, then it does not yield justified quasi-perceptual beliefs at all. Let us call this the Proper Constraint Dilemma:

The Proper Constraint Dilemma:

Background Assumption: The content of an act of imagination is constrained either properly or improperly.

Horn (I): If it is constrained properly, then the act of imagination may yield justified quasi-perceptual beliefs, but then the beliefs are justified in virtue of the the *proper constraints on* imagination, not in virtue of imagination itself. (3.14)

Horn (II): If it is constrained improperly, then the act of imagination does not yield justified quasi-perceptual beliefs.

Dilemma: In neither (I) nor (II), imagination is a source of justified quasi-perceptual beliefs.

Now, the above-mentioned concepts of *preservative* and *generative justification* (which I omitted from (3.14)) are explicitly defined by (Miyazono and Tooming, 2023a, §2–3), but these definitions are somewhat intricate, contentious, and hard to explain briefly; c.f. Lackey (2005); Egeland (2021); Miyazono and Tooming (2023b). Fortunately, the details of these notions are irrelevant for my present purpose. All that we need to understand for my purpose are the following similarities and differences between the Constraint Dilemma (3.12) put forward by Kinberg and Levy (2022) and the Proper Constraint Dilemma put forward by Miyazono and Toom-

ing (2023a), described above.

To begin, I note that Horn II of the Proper Constraint Dilemma is interesting because it elevates the Constraint Claim (3.11), which specified a *sufficient* condition for when imagination yields justified beliefs — if imagination is properly constrained, then it can yield justified beliefs — into a *necessary* condition: if imagination is *not* properly constrained, then it does *not* yield justified beliefs. I have already mentioned that this necessary condition is accepted by many, albeit contested by some; e.g. (Stuart, 2020; Gauker, 2021). Kinberg and Levy (2022) do not commit to this necessity claim; they limit the scope of their dilemma only to the sufficiency claim, as I have noted several times. Conveniently, however, Horn II of the Proper Constraint Dilemma is not the target of Miyazono and Tooming's discussion either, so I shall lay it aside.

Horn I of the Proper Constraint Dilemma (3.14) states that, *if* a constrained act of imagination yields a justified quasi-perceptual belief, then this belief is justified in virtue of the *constraints on* imagination, rather than in virtue of 'imagination itself'. The claim presented in this Horn is similar to Horn I of the Constraint Dilemma (3.12), but the two are not the same: Horn I of the Proper Constraint Dilemma is stronger than Horn I of the Constraint dilemma because the former concerns *all types* of constraints on imagination, both deliberate and indeliberate, rather than just deliberate constraints, as the latter does.

The target of Miyazono and Tooming's discussion is Horn I of the Proper Constraint Dilemma. Miyazono and Tooming (2023a) reject Horn I of this Dilemma on the basis of the idea that, *while performing an act of imagination, we can use our imagination to "tap into" constraints on imagination to which other cognitive faculties do not have access.* They call this claim INACCESSIBILITY, which I quote here for the sake of clarity (p.9):

> INACCESSIBILITY: At least in some cases of the epistemic use of imagination, imagination is properly constrained by some imaginative constrainers that are "cognitively inaccessible" in the sense that

> only imaginative processes can tap into the imaginative constrainers for the purpose of belief formation and that other belief-forming processes, such as perceptual or inferential processes, do not have access to them.

Call constraints on imagination that fall under the scope of the INACCESSIBILITY claim *'otherwise-inaccessible' constraints*: they are constraints that imagination can "tap into", but which are otherwise inaccessible (to other cognitive faculties). While Miyazono and Tooming (2023a) remain rather noncommittal about the *nature* of these constraints and intend to keep the scope of their INACCESSIBILITY claim as wide as possible, they do specify that with "imaginative constrainers" they mean "the prior representations that constrain the development of a scenario in imagination" (p.3) — but they also keep the door open for non-representational constraints, such as constraints grounded in the "cognitive architecture of imagination" (p.3, fn.1).

I wish to note that this wide conception of "imaginative constrainers" based on the idea of "prior representations" allows for some rather trivial cases where INACCESSIBILITY is true: for example, (i) cases where our imagination is constrained by *episodic memories* that have not been exhaustively interpreted propositionally; or (ii) cases where our imagination is constrained by episodic memories that are not immediately accessible to the imaginer, but which nonetheless constrain our acts of imagination. I provide a brief example of each case.

Recall Example 3 from Section 3.3.3: the example where you *use imagination* to count *in your episodic memory* how many windows there were in the house you grew up in. Suppose, for the sake of argument, that the house you grew up in and all records of the house have been destroyed, and you are the only person on earth with reliable episodic memories of the house's interior. In this case, your episodic memories of the windows in each room are examples of otherwise-inaccessible constraints on imagination: imagination can "tap into" these constraint provided by memory, but other cognitive faculties such as perception and reason cannot — rea-

son cannot "tap into" these constraints because it is not propositionally available, and perception cannot "tap into" these constraints because the house no longer exists. Perhaps the example is even more convincing if we suppose that some of these memories are not even straightforwardly (voluntarily) accessible by *memory itself* at any given time, but will only 'present themselves' to us if we *imagine* or remember an appropriate scenario that 'triggers' these memories for us (just like having a cup of tea with some Madeleines triggered a flood of vivid memories in Proust's *In Search of Lost Time*). This relates to the second case mentioned above, to which I turn next.

Consider the scene from the motion picture *Saving Private Ryan* where private Ryan tells captain Miller that he cannot visually remember his brothers' faces anymore:

> RYAN: I can't see my brothers' faces, man, and I've been trying and I can't see their faces at all. Is that ever happen to you?
> MILLER: You gotta think of a context.
> RYAN: Who's that?
> MILLER: We don't just think of other faces. Think about something specific, something you've done together. If I want to think of home, I think of something specific. I think of my hammock in the backyard, my wife pruning the rose bushes, my pair of old work gloves.

Presumably, Ryan could achieve the same by *imagining* his brothers' faces in some specific context (albeit perhaps less effective). This would also make the INACCESSIBILITY claim true.

Miyazono and Tooming (2023a) do not have such trivial examples in mind, however, as will transpire below.

I review one example of otherwise-inaccessible constraints, discussed by Miyazono and Tooming (2023a, pp.12–13), that is less trivial than otherwise-inaccessible episodic memories:

> INACCESSIBILITY is supported by the empirical studies of mental simulation, including Schwartz and Black's (1999) studies. In one of

> their experiments (Experiment 1), participants compared a narrow cup and a wide cup of the same height (both filled with water to the same height) and considered which cup needs a greater tilt before water spills from it (the correct answer: the narrow cup needs a greater tilt before water spills from it). Only a minority of participants gave the accurate answer in the condition where the question was asked verbally and descriptively (e.g., "Do you think the water pours out at the same or different angles for each cup?"). In contrast, their answers were more accurate in the condition where they held a cup, imagined it with their eyes closed, and tilted the cup until the water reached the rim in their imagination. A similar result was observed in another experiment (Experiment 3) in which participants simply visualized the cup without holding it.

Miyazono and Tooming (2023a) then go on to describe how particularly Experiment 3 of the study by Schwartz and Black (1999) is plausibly an example of INACCESSIBILITY:

> Suppose that Naomi, a participant in the visualizing condition (Experiment 3), came to the correct conclusion that the narrow cup needs a greater tilt. INACCESSIBILITY seems to be true about Naomi's case. First, it is very likely that Naomi's epistemic use of imagination was properly constrained given the fact that she came to the correct conclusion that the narrow cup needed a greater tilt. Second, it is also very likely that the relevant imaginative constrainers, in virtue of which Naomi came to the correct answer, were cognitively inaccessible. The imaginative constrainers were only available to imaginative processes, and other belief-forming processes did not have access to them, which is consistent with the finding that participants made systematic errors in non-visualizing conditions. Schwartz and Black insist that "people did not have a propositional knowledge base, explicit or implicit, that could have led to the accurate tilting" (Schwartz and Black, 1999, p.131), that "the intuitive knowledge found in simulated doing is not always on the same plane as other forms of knowledge", and that "it is not a trivial matter to connect epistemic planes that draw inferences using fundamentally different variables" (Schwartz and Black, 1999, p.134).

The constraints that imagination "taps into" here can be called *spatial* constraints: constraints on the spatial relations of our quasi-perceived scenario. Thus, the reason to think that the INACCESSIBILITY claim is true, is the same as the reason to think why Horn I of the Kinberg and Levy's Constraint Dilemma (3.12) is false; recall 3.4.4. (Miyazono and Tooming (2023a) note that this example quoted above does not, strictly speaking, *imply* that these constraints are otherwise-inaccessible, but it surely makes it *plausible*.)

I next emphasise that, strictly speaking, the INACCESSIBILITY claim by Miyazono and Tooming (2023a) is only a claim about constraints on imagination and our access to them, it is *not* a claim about the *reliability* of the quasi-perceptual beliefs gained via an act of imagination that is constrained by otherwise-inaccessible constraints. But there is no *prima facie* reason to believe that these beliefs are *never* reliable. The reliability of such beliefs must be evaluated on a case-by-case basis — or plugged in by assumption, as I regularly do with the condition that the relevant episodic memories (functioning as 'otherwise-inaccessible constraints', like in the trivial example discussed above) are reliable. Presumably, *expertise* will play a role in this evaluation; recall the Reliability Claim 3.13 from Section 3.4.5.

If such quasi-perceptual beliefs gained via an act of imagination that is constrained by otherwise-inaccessible constraints can be reliable, *then* these beliefs be justified. If also *true*, then these genuinely amount to knowledge. Thus the INACCESSIBILITY claim leads us to the claim that imagination can be a source of *knowledge* that is inaccessible to other types of knowledge-yielding processes, such as 'purely' perceptual or inferential processes. This claim is remarkable, but it is not new. It can be traced back at least to the account of thought-experiments from Mach (1897, 1960) and is often endorsed in relevant, recent literature. Gendler (1998, p.415, my italics), for example, writes:

> We have stores of unarticulated knowledge of the world which is not organized under any theoretical framework. *Argument will not give*

> *us access to that knowledge, because the knowledge is not propositionally available.* Framed properly, however, a thought experiment can *tap into it*, and — much like an ordinary experiment — allow us to make use of information about the world which was, in some sense, there all along, if only we had known how to systematize it into patterns of which we are able to make sense.

I shall return to this claim in Section 3.5.3, when I discuss sources of otherwise-inaccessible *knowledge*. I do wish to note at this point that the notion of "unarticulated knowledge" relates closely to practical knowledge, or knowledge-*how*: we *know how* to walk without falling, for example, and how to avoid bumping into others on a busy street, even though most of us have no *propositional* knowledge of how to walk — you just *do* it.

Enough about justification for now. I again repeat: if a justified quasi-perceptual belief (3.10) is *true*, then it is *quasi-perceptual knowledge*. We may now ask ourselves: what is the *source* of this knowledge, is it imagination, reason, or memory? The answer to this question will depend on what exactly we mean by a 'source of knowledge'. To this I turn now.

3.5 Sources of knowledge

The central research question of this Chapter is the Question of Knowledge Through Imagination: is imagination a source of knowledge of the natural world? In the previous Section, I made clear that I limit my attention to imagination as a source of a particular type of knowledge: *quasi-perceptual* knowledge. I explicated the concept of quasi-perception (3.7) and I elaborated at length about the processes of obtaining quasi-perceptual beliefs (3.9) and of their justification (3.10). In this Section, I turn my attention to the question what it means for something to be a *source of quasi-perceptual knowledge*. There is some ambiguity in the phrase "source of knowledge", so a clear understanding of the various ways in which something can be a source of knowledge will help us understand why it is often argued that imagination is *not* a source of knowledge of the world, even

though acts of imagination often *seem* to yield knowledge of the world; recall the three examples from Section 3.3.3.

To illustrate the pervasive ambiguity in the relevant literature, consider the following quote from Kinberg and Levy (2022, p.4), who directly argue against the idea that imagination is a source of knowledge. Kinberg and Levy (2022) often speak of a "special form of knowledge via the imagination" (p. 12) or "knowledge that is distinctively *from the imagination*" (p. 1), by which they mean that (p. 4):

> the issue at hand is not whether the imagination is *in some way or other involved* in knowledge acquisition. Rather, the question concerns the warrant-providing [i.e. justifying] role of the imagination. Or, to put this in less internalist-sounding terms: not any demonstration that (concrete, close-to-home) knowledge can be obtained partly by use of the imagination will do. What's at issue is whether, and if so when, the fact that a belief that p was produced via the imagination leads, in some distinctive way, to knowledge that p.

These remarks by Kinberg and Levy are highly ambiguous, specifically in absence of an explicit criterion for when a source of knowledge is 'distinctively imaginative'. (Recall Section 3.4.1, where I explained the justification of quasi-perceptual belief depends on the reliable functioning of no less than *three* cognitive faculties: imagination, memory and reason.) In this Section, I shall propose and discuss three explicit criteria for different 'types' of sources of knowledge that may be considered 'distinctively imaginative': (i) *basic* sources of knowledge (Section 3.5.1), (ii) *crucial* sources of knowledge (Section 3.5.2) and (iii) sources of *otherwise-inaccessible* knowledge (Section 3.5.3).

3.5.1 Basic sources of knowledge

When the phrase "sources of knowledge" is used in the literature, often the authors actually mean *basic sources of knowledge*. I follow Audi (2005) and understand a basic source of knowledge to be a source that can yield

knowledge *on its own*, in the sense that it "yields knowledge without positive dependence on the operation of some other source of knowledge (or justification)" (Audi, 2005, p.72).

I explicate this notion of a basic source of knowledge as follows. Consider a knowledge-yielding process $\mathcal{P} := \langle \mathcal{B}, \mathcal{J} \rangle$, consisting of a belief-yielding process $\mathcal{B}(q)$, yielding the true belief that q, and the justification of the obtained belief $\mathcal{J}$. Then:

> [Basic Source of Knowledge] Source of knowledge K is *basic* iff there exists a knowledge-yielding process $\mathcal{P} = \langle \mathcal{B}(q), \mathcal{J} \rangle$ such that both $\mathcal{B}(q)$ and $\mathcal{J}$ positively depend on the operation of *only* K. (3.15)

On the meaning of 'positive dependence', Audi (2005) remains somewhat unclear.[86] Fortunately, on the basis of my preceding discussion of processes of obtaining quasi-perceptual *beliefs* (Section 3.3), I can formulate the meaning of 'positive dependence' clearly, at least for the cases of quasi-perceptual knowledge: a quasi-perceptual knowledge-yielding process is positively dependent on the operation of sources of knowledge $K_1, ..., K_n$ iff all and only $K_1, ..., K_n$ are involved in the rational determination of the (true) quasi-perceptual belief (3.9) and in the process of its justification (3.4). Conveniently, this criterion can be simplified. I assumed that justification (3.4) resides only in the reliability and robustness of the belief-yielding process. Hence, justification will not involve *more* sources of knowledge than were already present in the process of belief-acquisition. Therefore, we need to look only at the process of belief-acquisition to find the sources of knowledge that the quasi-perceptual belief is 'positively dependent' on: a quasi-perceptual knowledge-yielding process is positively dependent on the operation of sources of knowledge $K_1, ..., K_n$ iff all and

[86] Positive dependence can be contrasted with negative dependence, which is dependence on *defeaters*. For example, observing a clock under normal conditions and perceiving that it reads 14:14 yields the perceptual knowledge that the current time is 14:14, which may be defeated — in this case: rendered unjustified and presumably false — if someone then comes by and tells you that the battery of your clock died yesterday.

only $K_1, ..., K_n$ are involved in the rational determination of the quasi-perceptual belief (3.9).

Before we can determine which sources of knowledge are *basic* sources, we need to know which sources of knowledge *exist* — what do the K's above denote? This is a highly controversial question largely beyond the scope of this Chapter, so rather than putting forward arguments for a purportedly exhaustive list of sources of knowledge, I will just put forward the five most probable and oft-discussed candidates, see e.g. (Hart, 1965; Lackey, 1999; Cohen, 2002; Audi, 2005; Leonard, 2023):[87]

- perception,
- reason,
- memory,
- imagination,
- testimony.

Of these five sources of knowledge, testimony is the odd one out because it is the only mentioned source that is not also considered to be a *cognitive faculty*. I include it nonetheless because it *is* an often-discussed source of knowledge, and because I wish to note that testimony, as a source of knowledge, may relate to memory and imagination in at least two interesting ways.

Firstly, memory and imagination are in some (possibly interesting) metaphorical sense sources of knowledge that function as *testimonials of*

[87] I note that Audi (2005) also mentions *consciousness*, a.k.a. *introspection* or *intuition*, as a "standard basic source of knowledge". In this Thesis, I lay aside consciousness because consciousness is a source of knowledge of our *inner world*, not of the external *natural world*, hence it is outside the scope of this Chapter. (I note moreover that consciousness problematises the notion of 'basic sources of knowledge', because even an uncontroversial basic source of knowledge such as ordinary perception is arguably positively dependent on consciousness of our perceptual *experience*; c.f. Audi (2005).) I note that imagination should not be regarded as an instance of introspection, even though it seems that imagination *is* an 'introspective' phenomenon. Sure, imagination requires 'introspective awareness' of one's mental state of imagination, but so does e.g. perception require 'introspective awareness' of one's mental state of perception.

past perceptions. It may be an interesting direction for future research (beyond the scope of this Thesis) to investigate this link between testimony and memory and imagination further.

Secondly and relatedly, knowledge gained by performing thought experiments (a type of epistemic act of imagination, see next Chapter) may be understood as a non-trivial combination of knowledge through imagination and through testimony, for the following reason. Thought experiments are often deliberately construed and communicated with the aim of conveying some specific insight, thus this insight is arguably gained partly through testimony (from the one who *constructed* and communicated the thought experiment) *and* partly through imagination (by the *performer* of the thought experiment). I return to this last point in Chapter 4, in which I analyse thought experiments specifically.

With these five sources of knowledge — perception, reason, memory, testimony and imagination — in hand, we can now look back at the two-step schema for quasi-perceptual beliefs (3.9) and recognise clearly that quasi-perception — memory or imagination — is *not* a basic source of knowledge. Recall:

The two-step schema for quasi-perceptual beliefs:
(1) On the basis of quasi-perceiving (concrete observable) entity ε, proposition q with topic ε comes to mind;
(2) On the basis of the *meta-belief* that the quasi-perceived scenario accurately represents the natural world, the propositional attitude of *belief* is adopted to q. (3.9)

This schema (3.9) involves *both* quasi-perception (step 1) *and* reason (step 2): the first step involves imagination or memory, or both (but nothing else), but the second step is an (implicit or explicit) *inference on the basis of a meta-belief* — it involves *reason*. So, neither imagination nor memory are basic sources of knowledge. In fact, the two-step schema for quasi-perceptual beliefs (3.9) shows us that *imagination and memory*

are not even basic sources of belief, let alone basic sources of knowledge.

This is as far as I venture into the topic of *basic* sources of knowledge. Imagination and memory are *not* basic sources of knowledge, full stop. I do not see much merit in further dissecting the notion of basic sources of knowledge.

In fact, I wish to note that many subtle issues may be brought up against the idea that there exist basic sources of knowledge at all. Consider the case of ordinary perception, which *is* a well-respected and uncontroversial basic source of knowledge (Audi, 2005). (This also follows from my criterion for basic sources of knowledge (3.15) and the two-step process for ordinary perception (3.8): the two-step process for ordinary perception mentions *only* ordinary perception, nothing else.) One may argue, however, as I noted in Section 3.2.2, p.112, that memory and imagination are in an important sense *essential* to perception; c.f. Audi (2005). This may be used to argue that perception does *positively depend* on memory and imagination as a source of knowledge (presumably *non-declarative* memory and *non-occurrent* imagination): without having perceptual concepts readily available in our memory, our perception would be rather 'blind'; and likewise, it has been argued, "most if not all perceptual experiences are infused with imagination" (Brown, 2018, p.133). I shall not linger on this issue, as my present purpose is to find out how (occurrent) imagination functions as a *distinctive* source of knowledge, not just how it is or is not a *basic* source of knowledge. I only wished to show that those who argue that imagination is *not* a basic source of knowledge, as e.g. Kinberg and Levy (2022) appear to do, are chasing a red herring.

In conclusion, it is safe to say that memory and imagination are *not* basic sources of knowledge (3.15). But to say that something is not a *basic* source of knowledge is not to say that it is not a source of knowledge at all. There are *other* ways in which imagination can function as a source of knowledge — not as a basic source, but at least as a *distinctive* source of knowledge. I turn to the second option I wish to take under consideration: *crucial* sources of knowledge.

3.5.2 Crucial sources of knowledge

With a *crucial* source of knowledge, I mean the following. Consider again a knowledge-yielding process $\mathcal{P} = \langle \mathcal{B}(p), \mathcal{J} \rangle$, consisting of a belief-yielding process $\mathcal{B}$, yielding the true belief that q, and the justification of the belief $\mathcal{J}$. Then:

> [Crucial Source of Knowledge] Source of knowledge K is *crucial* iff there exists a knowledge-yielding process $\mathcal{P} = \langle \mathcal{B}(q), \mathcal{J} \rangle$ such that both $\mathcal{B}(q)$ and $\mathcal{J}$ positively depend on the operation of K (amongst other K's). (3.16)

Thus, a source of knowledge is a *crucial* source of knowledge for a given knowledge-yielding process iff that source of knowledge *crucially*, i.e. irreducibly, contributes in the knowledge-yielding process to *both* the belief-forming process *and* to the justification of the belief. (If it is also the *only* source that contributes to the process, then it is a *basic* source of knowledge.) Although this idea is often implicitly adopted in the literature and it is rather straightforward, I am not aware of any explicit criteria for what makes a source of knowledge *crucial* for a knowledge-yielding process in the sense that I explicated in (3.16). Hence I proposed my own explication.

Before I continue, I make one elucidatory remark about the notion of a crucial source of knowledge (3.16). A crucial source of knowledge is a source of knowledge that positively contributes to every step of a given knowledge-yielding process. But this is not to say that the obtained knowledge — say, the knowledge that the current time is 14:14 (q) — can be obtained *only* via this source of knowledge. One may obtain the knowledge that q also via *other* knowledge-yielding processes. To illustrate this for the case of perception: you can *perceive* that the current time is 14:14, in which case perception is *crucial* for this process (and arguably even *basic*); but you can also come to know that the current time is 14:14 via e.g. reason or testimony. Crucial sources of knowledge need not have

'privileged access' to knowledge — see next Section for more on this.

Is imagination a crucial source of quasi-perceptual knowledge (3.16)? It is evident that the *formation* of (true) quasi-perceptual beliefs (3.9) positively depends on imagination. What about their justification? Well, as I argued in Section 3.4, there are at least *three* cases in which imagination can be a source of *justification* for quasi-perceptual beliefs, even in the face of the Constraint Dilemma (3.12) and the Proper Constraint Dilemma (3.14).

(i) The first case where imagination can be a source of justification for quasi-perceptual beliefs is when an act of imagination is deliberately constrained and the act of imagination yields a quasi-perceptual belief through a process where the *spatial properties* of the quasi-perceived scenario are epistemically crucial (Section 3.4.4). The clearest examples here were the cases where you learn, by *crucially* using the imagination, how many elephants would fit in your room (Example 1 from Section 3.3.3) or whether the poem that you wrote for your best friend's wedding fits to Queen's *We Will Rock You* (Example 2 from Section 3.3.3).

(ii) The second case where imagination can be a source of justification for quasi-perceptual beliefs is when an act of imagination is indeliberately constrained and the act of imagination concerns a context wherein the imaginer is an *expert*, in which case the indeliberate constraints on imagination can nonetheless yield *reliable* quasi-perceptual beliefs (Section 3.4.5). Notable examples here were e.g. car mechanics who can easily answer correctly that heating up a metal disk with a hole in the middle will *increase* rather than decrease the size of the middle hole (Figure 3.6), and, in general, scientists performing thought experiments in domains where they are experts; see (Kuhn, 1977) for extensive discussion.

(iii) The third case is when an act of imagination is constrained by constraints that are *accessible* only to imagination and *inaccessible* to other sources of knowledge and justification. These constraints will often concern episodic memories or spatial properties of a quasi-perceived scenario (Section 3.4.7). (If these otherwise-inaccessible constraints are *deliberate*

constraints, then this way in which imagination can be a crucial source of knowledge collapses to the first case, described above. But I do not believe that it is necessarily the case; nor do I believe that it is necessarily the case that otherwise-inaccessible constraints are always *indeliberate* constraints. So we should keep the two cases distinct.) A clear example of this case was put forward by Miyazono and Tooming (2023a), recall Section 3.4.7: participants in the empirical studies of Schwartz and Black (1999) had to compare a narrow cup and a wide cup of the same height and consider which cup needs a greater tilt before water spills from it — this answer was answered incorrectly more often than not when imagination was *not* employed, and it was answered correctly more often than not when imagination *was* employed.

I conclude that imagination is a crucial source of knowledge (3.16).

3.5.3 Sources of otherwise-inaccessible knowledge

Finally, I discuss sources of *otherwise-inaccessible* knowledge. Recall Section 3.4.7, where I discussed how it is occasionally argued that imagination can yield knowledge that no other source of knowledge has access to. This claim was made plausible by the INACCESSIBILITY claim put forward and defended by Miyazono and Tooming (2023b). The claim can be traced back to the work of Mach (1897, 1960), and has a notable contemporary proponent in Gendler (2004), amongst others. Recall e.g. that Gendler (1998, p.415, my italics) wrote:

> We have stores of unarticulated knowledge of the world which is not organized under any theoretical framework. *Argument will not give us access to that knowledge, because the knowledge is not propositionally available.* Framed properly, however, a thought experiment [and, presumably, other types of acts of imagination] can *tap into it*, and — much like an ordinary experiment — allow us to make use of information about the world which was, in some sense, there all along, if only we had known how to systematize it into patterns of which we are able to make sense.

Now, I flag that we should be careful how we interpret Gendler's claim. Gendler claims that we have "stores of unarticulated knowledge" *stored within ourselves*, which can be accessed only through imagination. I noted in Section 3.4.7, that this notion of "unarticulated knowledge" is closely related to *practical knowledge*, or know-how, as in e.g. most of us know *how* to walk but have little to no propositional knowledge about this — imagination can then "provide access" to this propositional knowledge. A more concrete example (the case of water spilling out of a tilted tall-small or short-wide glass) was discussed by Miyazono and Tooming (2023a) in support of their INACCESSIBILITY claim; again, recall Section 3.4.7. In these cases, imagination can function as a *crucial* source of knowledge (3.16), as I argued in the previous Section. But, even though using imagination may be the only way to obtain such knowledge 'from within ourselves', i.e. *without novel external (empirical) input*, it may very well be the case that such knowledge may be obtained via *other* means that *do* involve external input. We may straightforwardly obtain such knowledge through perception, for example, or through testimony.

Nonetheless, Miyazono and Tooming's INACCESSIBILITY claim and Gendler's above-quoted claim raise the question whether there are tokens of knowledge that can *only* be obtained by using the imagination, and not through any other means. This would mean that imagination is, what I call, a *source of otherwise-inaccessible knowledge.* I explicate this notion of a source of otherwise-inaccessible knowledge as follows:

> [Source of Otherwise-Inaccessible Knowledge] Source K is a source of *otherwise-inaccessible* knowledge iff there exists a proposition q such that K is *crucial* (3.16) for *all* knowledge-yielding processes that yield the knowledge that q. (3.17)

This notion of otherwise-inaccessible knowledge seems rather arcane, but I would argue that it is surprisingly commonplace. Intuitively, most of our sources of knowledge *are* sources of otherwise-inaccessible knowledge.

To illustrate: reason is arguably a source of otherwise-inaccessible knowledge of many types of inferential knowledge (e.g. that there are infinitely many primes); memory is often a source of otherwise-inaccessible knowledge about the past (e.g. how many times you just saw a fly crashing into the window a minute ago); and even testimony can be considered a source of otherwise-inaccessible knowledge of propositions which we cannot come to know *on our own*, such as complex scientific knowledge that we cannot generate on our own but which instead must be *taught* to us.[88]

Is imagination a source of otherwise-inaccessible knowledge? Yes it is, given that (i) the INACCESSIBILITY claim put forward by Miyazono and Tooming (2023a) is true, and (ii) a true quasi-perceptual belief is obtained via an act of imagination constrained by such otherwise-inaccessible constraints are reliable, i.e. justified, *and* (iii) the yielded knowledge is *not accessible otherwise.* If (i) and (ii) are the case, then imagination was a *crucial* source of knowledge (3.16). For imagination to additionally be a source of *otherwise-inaccessible* knowledge (3.17), it must be the case that imagination is a crucial source of knowledge for some token of knowledge that *cannot be obtained through other means than through imagination.*

One Example that comes close to being a case where imagination is a source of otherwise-inaccessible knowledge was the variation of Example 3 (Section 3.3.3) that I discussed in Section 3.4.7. Recall Example 3 from Section 3.3.3, which was the Example where you *use imagination* to count *in your episodic memory* how many windows there were in the house you grew up in. We supposed next, for the sake of argument, that the house you grew up in and all records of the house have been destroyed, and you are the only person on earth with reliable episodic memories of the house's interior. In this case, your episodic memories of the windows in each room are examples of otherwise-inaccessible constraints on imagination: imagination can "tap into" these constraints provided by memory, but other

[88] The big star of sources of otherwise-inaccessible knowledge is, of course, *introspection*, which has an *entire domain* of knowledge accessible to only itself: self-knowledge. But I laid this source of knowledge aside in Section 3.15, as it gives knowledge of our *inner world*, not of the natural world; recall footnote 87, p. 185.

cognitive faculties such as perception and reason cannot — reason cannot "tap into" these constraints because it is not propositionally available, and perception cannot "tap into" these constraints because the house no longer exists. If it is the case that *memory*, 'by itself', cannot "tap into" these constraints either but that you really need to engage with these memories and "tap into them" *in the imagination*, then we have here an example where imagination is a source of otherwise-inaccessible knowledge.

Of course, in this example, we must be careful not to confuse the contribution of *imagination* with the contribution of *memory*, in obtaining this otherwise-inaccessible knowledge. It must be carefully evaluated whether imagination is a *crucial* source of knowledge in every case — alongside the contribution of memory. I suspect that it will be a persistent problem to disentangle the contribution of *imagination* from the contribution of *memory* to obtaining otherwise-inaccessible knowledge in most cases that may present themselves as possible examples of imagination as a source of otherwise-inaccessible knowledge.

Beyond this Example (and other variations of the same type), I am currently unaware of clear examples of a different type where imagination is a source of *otherwise-inaccessible* knowledge (3.17), but I can see no principled reason for why this could *not* be the case. I optimistically conclude, then, that imagination *is* a source of otherwise-inaccessible knowledge.

3.6 Conclusion

To conclude and recapitulate, in this Chapter, I discussed the Question of Knowledge Through Imagination: is imagination a source of knowledge of the natural world?

I began by explicating the notion of *quasi-perception* (3.7), which denotes both *acts of imagination* (3.1) and *episodic memories* (3.5). I discussed at length the similarities and differences between quasi-perception and 'ordinary' perception. Following Dorsch (2016b), in analogy to a two-step reconstruction of how 'ordinary' perceptual beliefs are formed(3.8),

I then put forward a *two-step schema* for the rational determination of quasi-perceptual beliefs (3.9). To motivate this two-step schema, I discussed three examples where quasi-perceptual beliefs are obtained (Section 3.3.3) and I argued why the formation of a quasi-perceptual belief necessarily requires *meta-beliefs* about the accuracy of our quasi-perceptions, which are involved in step 2 in the two-step schema for quasi-perceptual beliefs (Section 3.3.4).

I then discussed how quasi-perceptual beliefs are justified. I first provided an explicit criterion for the justification of quasi-perceptual beliefs (3.10). I then discussed the Constraint Claim (3.11), which is the widely-endorsed claim that imagination can be a source of reliable and robust, i.e. *justified*, quasi-perceptual beliefs if the content of our imagination is *properly constrained in a reality-oriented way*.

The Constraint Claim was challenged by Kinberg and Levy (2022), who argued that it gives rise to a dilemma (3.12). The dilemma ran as follows. The content of an act of imagination is either (I) deliberately constrained or (II) indeliberately constrained. Horn (I): if an act of imagination is deliberately constrained, then it may yield justified quasi-perceptual beliefs, but the beliefs are not justified *in virtue of imagination*. Horn (II): if an act of imagination is indeliberately constrained, then it never yields justified quasi-perceptual beliefs. I argued first that the literature on scientific thought experiments has taught us that Horn (I) is false (Section 3.4.4, see also next Chapter). I next argued that Horn (II) is also false, because an indeliberately-constrained act of imagination can yield justified beliefs if the imaginer is an *expert* in the imagined topic; recall the Reliability Claim (3.13). Finally, I argued that the dilemma put forward by Kinberg and Levy (2022) is a false dilemma, as our acts of imagination are typically constrained by non-trivial *combinations* of deliberate and indeliberate constraints, which may interact in epistemologically interesting — and epistemically valuable — ways (Section 3.4.6). Finally, I reviewed a similar dilemma (3.14) discussed — and argued against — by Miyazono and Tooming (2023a). I concluded that imagination can indeed contribute

to the justification of quasi-perceptual beliefs.

Finally, I discussed what it means to say that "imagination is a source of knowledge". I distinguished three ways in which imagination may function as a source of knowledge: (i) as a *basic* source of knowledge (3.15), (ii) as a *crucial* source of knowledge (3.16), and (iii) as source of *otherwise-inaccessible* of knowledge (3.17). I concluded (i) that imagination is *not* a basic source of knowledge (Section 3.5.1), (ii) that imagination *is* a crucial source of knowledge in at least three different ways (Section 3.5.2), and (iii) that imagination is even a source of otherwise-inaccessible knowledge (Section 3.5.3).

Enough about the Question of Knowledge Through Imagination. I turn to the final main topic of this Thesis: scientific thought experiments.

Chapter 4

Scientific Thought Experiments

We talked about everything, ranging from the tree of life to the pituitary gland. Most of my knowledge was intuitive. I had a flexible imagination and was always ready for a game that we would play. Harry would test me with a question. The answer had to be a sliver of knowledge expanding into a lie composed of facts.

Patti Smith, *Just Kids*

4.1 Introduction

Scientific Thought Experiments (STEs): what do they do and how do they do it? Galilei dropped balls off the tower of Pisa, Newton rotated a bucket of water in an empty universe, Maxwell conjured up a Demon, and Einstein rode on light beams and in space-bound elevators. Thought experiments are everywhere, both in the history of science and in contemporary science, often with far-reaching consequences. Initiated by the likes of Mach, Koyré, Popper and Kuhn, and with exponentially increasing effort since the 1990s, philosophers of science have set out to explain the ubiquitous presence and seemingly revolutionary capabilities of STEs.

STEs first and foremost demand explanation because we often gain scientific knowledge and understanding by performing STEs, but STEs are performed in the *imagination*. The idea that we can gain scientific knowledge and understanding *just* by using our imagination is highly controversial. So, if STEs are performed in the imagination, then how, if ever, could we gain scientific knowledge and understanding by performing them?

A wide variety of accounts of STEs have been proposed in the literature, but there is still little consensus about two *core questions*:

(I) What *are* scientific thought experiments?

(II) What, and how, do we *learn* by performing scientific thought experiments?

In this Chapter, I propose a novel account of STEs explicitly based on the recently developed *fiction view of models*.[89] The fiction view of models construes scientific models (literally) as works of fiction, and it describes model-based reasoning as crucially *imaginative* engagement with a specific type of fictional worlds: model systems. I shall argue that these insights are directly relevant for understanding STEs. The main idea underlying the account of STEs that I shall propose, *the fiction view of STEs*, can be summarised as follows:

> **The fiction view of STEs:** To perform a scientific thought experiment is to reason with and about scientific models — construed as works of fiction — with an epistemic aim.

I shall argue that this account improves on similar recent proposals in the literature, notably (Meynell, 2014) and (Sartori, 2023), and that we can use it to provide illuminating answers to core questions (I) and (II) mentioned above. I proceed as follows.

[89] See e.g. Godfrey-Smith (2007); Frigg (2010a); Frigg and Nguyen (2016, 2018, 2017a, 2021b, 2020); Levy (2012, 2015); Toon (2012); Weisberg (2013); Salis and Frigg (2020); Salis (2016, 2021, 2020); Levy and Godfrey-Smith (2020). See Section 4.4.2 in this Chapter for discussion.

To begin, in Section 4.2, I elaborate on the *core questions* (I) and (II) posed above, and I sketch the current state of the debate concerning these questions (§4.2.1). I then introduce two example STEs — Galilei's falling bodies and Clement's Sisyphus — which I will analyse extensively throughout this Chapter, after which I briefly introduce four more STEs, each of which illustrate important characteristics of STEs (§4.2.2). I then return to core questions (I) and (II) and I provide a *definition* of what an STE is (4.3) and an *explication* of what it means to perform an STE (4.4) (§§4.2.3–4.2.4). I then discuss and evaluate the two main accounts of STEs available in the literature — the argument view and the mental-modeling view — indicating their respective strengths and weaknesses (§4.2.5).

In Section 4.3, I introduce and discuss at length the concept of *fiction*. I begin by elaborating on the relation between STEs and the concept of fiction (§4.3.1). I then introduce and discuss the theory of fiction that my proposed account of STEs is built upon: Walton's (1990) theory of fiction as make-believe (§4.3.2). I then discuss at-length two recently proposed accounts of STEs — the proposals by Meynell (2014) and Sartori (2023) — both of which employ Walton's theory of fiction. I indicate the advantages of these accounts and I indicate where these accounts must be improved.

In Section 4.4, I begin by discussing the relation between STEs and scientific models (§4.4.1). I then introduce the *fiction view of models* (§4.4.2), after which I argue that this philosophical account of scientific modeling is directly relevant for understanding STEs (§4.4.3).

In Section 4.5, I introduce my proposed account of STEs: *the fiction view of STEs*. I begin by formulating the proposed account as clearly as possible (§4.5.1). I then return to the two main example-STEs introduced in Section 4.2 and analyse them at length using the proposed account (§§4.5.2–4.5.3); and I briefly do the same for the other four examples introduced in Section 4.2 (§4.5.4). I then discuss how my proposed account increases our understanding of STEs in ways that existing accounts of STEs have not (§4.5.5).

4.2 Scientific Thought Experiments

4.2.1 Core questions concerning STEs

Scientific thought experiments (STEs) have by now been philosophically scrutinised for over three decades. Countless accounts of STEs have been proposed in the literature. These accounts aim to provide answers to what I call the two *core questions* concerning STEs:

(I) What *are* STEs?

(II) What, and how, do we *learn* by performing STEs?

I briefly sketch the state of the debate concerning these core questions.

There is no shortage of answers to question (I), as STEs have been described as being intrinsically related to a wide variety of scientific activities, notably:

- ★ 'real' experimenting; e.g. (Gooding, 1992, 1993, 1994; Sorensen, 1992; De Mey, 2003; Arcangeli, 2018; Sartori, 2023),
- ★ argumenting; e.g. (Norton, 1993, 1996, 2004a,b; Häggqvist, 1998, 2009; El Skaf, 2018, 2021; Mulder and Muller, 2023),
- ★ computer simulation; e.g. (El Skaf and Imbert, 2013; Arcangeli, 2018; Lenhard, 2018; Murphy, 2020; Shinod, 2021),
- ★ scientific modeling and model-based reasoning; e.g. (Boniolo, 1997; Morgan, 2002, 2004; Cooper, 2005; Markie, 2005; Arcangeli, 2017; Salis and Frigg, 2020; El Skaf and Stuart, 2023), and
- ★ mental-modeling and other forms of analogical reasoning; e.g. (McMullin, 1985; Miščević, 1992, 2007; Nersessian, 1993, 1999, 2018; Gendler, 2004; Cooper, 2005; Smith, 2007; McAllister, 2012; Camilleri, 2014a; Clement, 2009a,b, 2018; Kornberger and Mantere, 2020).

As a consequence of this plurality of answers to the question what STEs *are*, there is also a plurality of answers to the question (II) what, and how,

we *learn* by performing them. Indeed, it has been argued that STEs help us achieve nearly all scientifically relevant epistemic aims, notably:[90]

- ✯ generating scientific knowledge (Brown, 1991; Norton, 2004b; Nersessian, 2018; Miščević, 2022),
- ✯ generating scientific understanding (Brown, 2014; Murphy, 2020; Stuart, 2016, 2018),
- ✯ instigating conceptual change (Kuhn, 1977; Nersessian, 1993, 1999, 2018; Steier and Kersting, 2019),
- ✯ constructing models (Boniolo, 1997; Morgan, 2002, 2004; Cooper, 2005; Markie, 2005),
- ✯ showing inconsistencies within a theory (Brown, 1991, pp.34-36),
- ✯ making a theory appear plausible (Brown, 1991, pp.36-38),
- ✯ clarifying our 'conceptual apparatus' (Kuhn, 1977),
- ✯ justifying an existing human made-system (Reiss, 2012),
- ✯ revealing background assumptions (Gendler, 1998) or questioning them (Camilleri, 2014a),
- ✯ proposing new theoretical possibilities (Stuart, 2021; El Skaf, 2021),
- ✯ revealing and resolving inconsistencies (El Skaf, 2021; El Skaf and Palacios, 2022; Sorensen, 1992; Häggqvist, 2009; Häggqvist, 2019),
- ✯ giving examples or illustrating a claim (Brown, 1991; Schabas, 2017),
- ✯ demonstrating pursuitworthiness (Miller, 2002; El Skaf, 2021),
- ✯ giving "hypothetical explanations" (Schlaepfer and Weber, 2017),
- ✯ controlling variables (Sorensen, 1992),
- ✯ exemplifying features (Elgin, 2014), and
- ✯ testing non-empirical virtues of models and theories (Bokulich, 2001).

[90] The first half of this list includes all epistemic aims mentioned by (Meynell, 2014, p.4162), the second half includes those mentioned by (El Skaf and Stuart, 2023, p.12).

In the face of these lengthy lists, one may wonder whether there is anything left to *explain* about STEs that has not been explained yet. I think there is. The many different accounts of STEs available in the literature are so disparate, conflicting and even *prima facie* incompatible, that there is plenty room for improvement, if only by providing an account of STEs that can *incorporate* and *harmonise* many of the *prima facie* conflicting insights provided by the accounts of STEs that are available in the literature. This shall be my primary aim: to provide an account of STEs that can explain, in one fell swoop, everything that has already been explained about STEs. My secondary aims with my proposed account, then, is to *increase* our understanding of STEs in ways that have not been noted before, and to indicate possibilities for fruitful future research.

I note that the scope of my analysis is limited to *scientific* thought experiments. I lay aside thought experiments performed in other disciplines, notably *philosophical* thought experiments. For the sake of clarity, I next distinguish, as clearly as possible at this stage, *scientific* TEs from non-scientific TEs. This distinction will be only provisional — I propose an explicit *criterion* for when a thought experiment is a *scientific* thought experiment in Section 4.5, in the context of my proposed account of STEs. (To get ahead of myself: a TE is a *scientific TE* only if the description of that TE is, or includes, the description of a scientific model.)

There is a branch of literature that focuses predominantly on thought experiments in *science*; e.g. Brown (1991); Norton (1996); Nersessian (1993); and a branch of literature that focuses predominantly on 'non-scientific', i.e. philosophical, TEs; see e.g. Häggqvist (1998); Cohnitz and Häggqvist (2017) and the references therein; and there is a branch of literature that focuses on *all* types of TEs and explicitly argues that it is wrong to make a principled distinction between scientific and non-scientific TEs; e.g. Davies (2007); Meynell (2014). I belong to the first branch and focus exclusively on scientific TEs in this Chapter. Although I agree that thought experiments across the sciences and the various domains of philosophy share important characteristics — all TEs are deliberately designed

and performed epistemic acts of imagination — I do believe that there is an important difference between scientific TEs and non-scientific TEs. Namely: that they have different types of *epistemic aims*.

Thought experiments have epistemic aims, and the epistemic aim of a given TE will generally be identifiable "within a specifiable problem domain" (Yeates, 2004, p.150), in the sense that it should be reasonably clear to which subdomain of science or philosophy a thought experiment belongs on the basis of its epistemic aim; c.f. (Brown and Fehige, 2019, §2). For example, Newton's bucket (see Section 4.2.2) conveys an argument pertaining to the definition of relative motion and, hence, belongs to (proto-)physics; Thomson's violinist (Thomson, 2004) conveys an argument in favor of abortion and, hence, belongs to ethics. So, we can provisionally distinguish scientific thought experiments from non-scientific thought experiments by limiting our attention to thought experiments that have an epistemic aim that is *relevant for science*.

I take it that, on the basis of this distinction, a distinction arises between 'obvious' scientific TEs like Galilei's falling bodies, Clement's Sisyphus, Newton's bucket, Maxwell's demon, Einstein's photon-box and Norton's dome and 'obvious' non-scientific TEs such as Thomson's violinist, Foot's trolley problem, Gettier problems, Putnam's Twin Earth and Searle's Chinese room. Several 'grey area' TEs will be discussed in Section 4.5. Until then, I focus my attention on the 'obvious' scientific TEs mentioned above.

I next introduce two examples of STEs that I often refer to throughout this Chapter and which I shall analyse extensively in Sections 4.5.2 and 4.5.3: Galilei's falling bodies and Clement's Sisyphus. I then briefly discuss a handful of other STEs (Newton's bucket, Maxwell's demon, Einstein's photon-box, and Norton's dome) that I shall occasionally refer to throughout this Chapter. I shall then return to the two *core questions* concerning STEs, where I use these examples to discuss the 'key components' and core characteristics common to all STEs, and to specify the key *desiderata* that an account of STEs must meet.

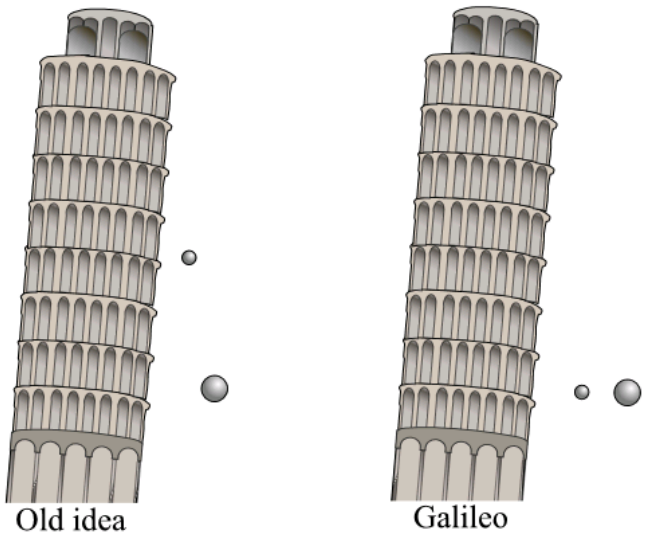

Figure 4.1: Galilei's falling bodies. (Image courtesy of wikipedia.com.)

4.2.2 Two examples — and four more

Example 1: Galilei's falling bodies

Perhaps the most famous and most extensively analysed STE in the literature is the falling bodies thought experiment presented by Galileo Galilei in his *Dialogues Concerning Two New Sciences*; see e.g. Brown (1986, 1991, 2004); Miščević (1992); Nersessian (1993); Norton (1993, 1996, 2004b); McAllister (1996); Gendler (1998); Reiner (1998); Arthur (1999); Lattery (2001); Peijnenburg and Atkinson (2003); Brendel (2004); Palmieri (2005); Cooper (2005); Norton and Roberts (2010); Palmerino (2012, 2018); Meynell (2014); Camilleri (2015); Aldea (2019); Brown and Fehige (2019); Murphy (2020); Gruszczyński (2022).

With his falling bodies thought experiment, Galilei argued against the Aristotelian principle that objects fall with a speed proportional to their mass, i.e. that every object has a 'natural speed' of falling that depends on the object's mass, such that heavy objects fall *faster* than light objects. Galilei argued against this principle by showing, by means of a "short and

conclusive" illustrative argument, that this principle leads to a contradiction. I quote the core part of Galilei's original presentation of this STE at length (Galilei, 1638, pp.62–63):

> But, even without further experiment, it is possible to prove clearly, be means of a short and conclusive argument, that a heavier body does not move more rapidly than a lighter one provided both bodies are of the same material and in short such as those mentioned by Aristotle. But tell me, Simplicio, whether you admit that each falling body acquires a definite speed fixed by nature, a velocity which cannot be increased or diminished except by the use of force [*violenza*] or resistance. [...] If we then take two bodies whose natural speeds are different, it is clear that on uniting the two, the more rapid one will be partly retarded by the slower, and the slower will be somewhat hastened by the swifter. [...] But if this is true, and if a large stone moves with a speed of, say, eight while a smaller moves with a speed of four, then when they are united, the system will move with a speed less than eight; but the two stones when tied together make a stone larger than that which before moves with a speed of eight. Hence the heavier body moves with less speed than the lighter; an effect which is contrary to your supposition. Thus you see how, from your assumption that the heavier body moves more rapidly than the lighter one, I infer that the heavier body moves more slowly.

Thus we have arrived at a contradiction. Assuming that objects fall at speeds propositional to their mass, when we tie two objects of unequal mass together, we can infer that the two objects, when joined together, must fall slower than the heaviest object, *and* that they must fall faster than the heaviest object. Galilei concludes: *the speed of falling objects is independent of their mass.*

I make two inter-related comments about this STE.

Firstly, Galilei's presentation of the falling bodies STE quoted above does at least two things: (i) it presents an *argument*, specifically a *reductio ad absurdum*, and (ii) it describes and invites us to *imagine* a rather specific imaginary scenario, i.e. a scenario where we tie together and drop

two stones of unequal weight (a 'large stone' and a 'smaller stone'). The epistemological question that has been debated endlessly in the literature is whether all the epistemic value of this STE must be found *only* in the argument that it presents (i), or whether there is also epistemic value to be found in the part of the STE that appeals to the *imagination* (ii). As will become clear in the next Sections, I side with the latter camp.

Secondly and relatedly, it is not obvious how Galilei's conclusion — the speed of falling objects is independent of their mass — *follows from* the argument presented in the above-quoted passage. Indeed, as we shall see in Section 4.5.2, extra background assumptions are required and some extra reasoning must be done, before we can *deductively arrive* at Galilei's conclusion that the speed of falling objects is independent of their mass.[91] This work will at least partially be done *in the imagination*, and so it must be evaluated carefully whether it is the case that imagination performs a crucial, i.e. irreducible, epistemic role in STEs, which harks back to the first comment above.

Example 2: Clement's Sisyphus

The second example STE that I wish to introduce is Clement's Sisyphus. This is a contemporary STE constructed by the cognitive scientist John J. Clement in order to empirically investigate the role of imagistic processes in analogical reasoning; see notably (Clement, 1998, 2009a,b, 2018). It has not been extensively analysed in the literature, but I chose it nonetheless because it highlights some important features of STEs that remain underilluminated by Galilei's falling bodies and the analyses thereof.[92]

[91] Which is only correct given the additional assumption – which Galilei (rather understandably) overlooked – that the center of gravity of the falling bodies have equal distance to the center of gravity of the earth; see e.g. (Atkinson and Peijnenburg, 2004, Appendix C) for discussion. [92] We are definitely concerned with an STE here: Clement (2009a) uses cases similar to the present example to *define* thought experiments as "the act of considering an untested, concrete system designed to help evaluate a scientific concept, model, or theory—and attempting to predict aspects of the system's behavior." This definition is very similar to the one proposed in this Chapter, without the explicit reference to the fiction view of models.

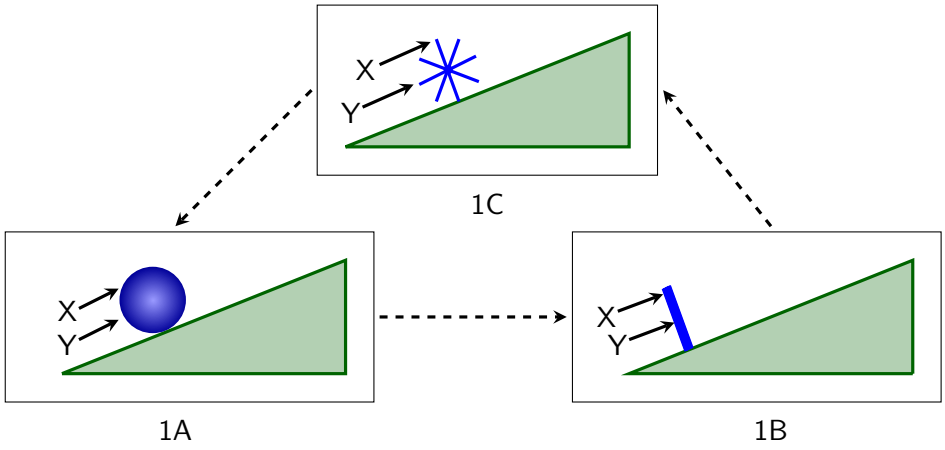

Figure 4.2: Clement's Sisyphus. Based on (Clement, 2009a, Figure 1), original description: "Analogies for Sisyphus problem."

Clement (2009a) presented several test-subjects (not physicists) with diagram 1A of Figure 4.2 and the following description:

> You are given the task of rolling a heavy wheel up a hill. Does it take more, less, or the same amount of force to roll the wheel when you push at X, rather than at Y? Assume that you apply a force parallel to the slope at one of the two points shown, and that there are no problems with positioning or gripping the wheel. Assume that the wheel can be rolled without slipping by pushing it at either point.

The test-subjects were instructed to 'think out loud' and indicate their reasoning processes by means of verbal and non-verbal communication (hand-movements, etc.). Diagrams 1B and 1C in Figure 4.2 represent the analogies that one test-subject, call her Alice, made in finding the solution to this thought experiment: first, Alice imagined a physical system with which she has real-life experience — lever system (1B) — and she was able to formulate an answer to Clement's question in this context: it is easier to push over the lever at X than it is at Y. Then, Alice convinced herself that you can make up a wheel by superimposing many levers (1C) such that the answer to the question remains the same, thus justifying the

intuition that the answer to Clement's question in the context of Diagram 1B carries over to the context of Diagram 1A.

I again make two inter-related comments about this STE.

Firstly, whereas it is an open question whether imagination plays a crucial epistemic role in Galilei's falling bodies, it is undeniable that imagination plays a crucial epistemic role in Clement's Sisyphus. The analogical reasoning performed by Alice is *not* exclusively propositional (it is not 'mere' reasoning through an argument), but also *crucially imagistic*.

Secondly, the act of imagination that is performed in Clement's Sisyphus STE is much more *improvised* and *open-ended* than the act of imagination performed in Galilei's falling bodies. Galilei's falling bodies leaves little room for variation, it *determines* much of the content of our imagination by explicitly presenting an argument, while Clement's Sisyphus leaves plenty room for variation. (Indeed, different test-subjects made different analogies when performing Clement's Sisyphus, see Section 4.5.3 in this Chapter.) But the act of imagination in Clement's Sisyphus is not entirely *free*: clear instructions have been given, such that the act of imagination is explicitly *constrained* by the background assumptions that the performer of the STE should adhere to (i.e. that you apply a force parallel to the slope at one of the two points shown, that there are no problems with positioning or gripping the wheel, and that the wheel can be rolled without slipping by pushing it at either point). Just like Galileo, Clement prescribes us to imagine a *specific imaginary scenario*; but unlike Galileo, Clement gives us the freedom to reason about this specific imaginary scenario however we want. And yet both are thought experiments.

Four more examples

I next introduce four famous STEs that I shall occasionally refer to throughout this thesis: Newton's bucket, Maxwell's demon, Einstein's photon-box, and Norton's dome. I note that all four of these STEs have many conflicting formulations and interpretations in the literature. I will not be able to do justice to all this variance. I chose to introduce these STEs in a form

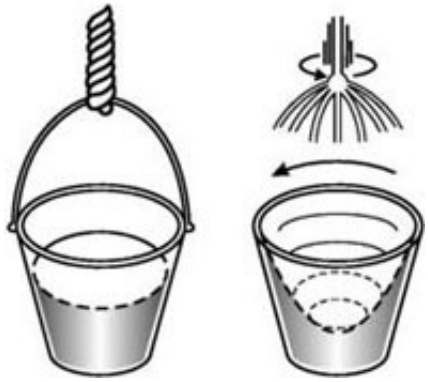

Figure 4.3: Newton's bucket. (Image courtesy of einstein.stanford.edu/SPACETIME/spacetime1.html.)

that is both short and still instructive to my purpose at hand: illustrating core characteristics of STEs.

(a) Newton's bucket. In a Scholium to Book I of the *Principia*, Isaac Newton presented the bucket thought experiment to argue against the Cartesian principle that motion can be defined as *relative* motion with respect to the object's immediate neighbouring objects.

Newton (1999) asked us to imagine a bucket filled with water attached to a rope that is twisted many times so that, when the bucket is released, the rope will unwind and the bucket will spin; see Figure 4.3. After the bucket has spinned for a while, the water will have risen up the sides of the inside of the bucket and, thus, the surface of the water has become concave. Newton next makes the following observation: the water that has risen up the sides of the inside of the bucket clearly has *rotational motion*, but it does not have *relative motion* with respect to the inside of the bucket, which also rotates, at the exact same speed. Hence, Newton argued, the Cartesian principle that motion can be defined as *relative* motion with respect to the object's immediate neighbouring objects is false.

It is important to note that in this STE Newton instructs us to imagine *only* the bucket of water and does not instruct us to imagine anything

else in its surroundings. This is why it can be said that the water in the rotating bucket does not have relative motion to *anything*. As such, it is often said that Newton thus implicitly prescribes us to imagine the bucket in *otherwise empty space*; see e.g. Winterbourne (1985); Barbour and Pfister (1995); Brown (2013). This crucial background assumption for the bucket thought experiment was famously objected to by Ernst Mach (which objection, it has been argued, was already anticipated by Leibniz in the famous Leibniz-Clarke debates); c.f. Bouquiaux (2008). Mach argued that the water in the bucket *does* have relative motion, namely relative motion *with respect to all the stars in the universe*. Mach thus refused to imagine a bucket in an otherwise-empty universe, because he rejected the assumption that water in such a bucket would behave just like water in a bucket rotating in our universe filled with mass. Further details of this debate are beyond the scope of this Chapter. I only presented Newton's bucket to show that STEs not only present *impossible scenarios* (such as spinning buckets of water in otherwise-empty universes) but also often *leave crucial assumptions implicit*, thus allowing for *conflicting interpretations* of one and the same STE. The objections to Newton's bucket by Ernst Mach illustrate this point.

(2) Maxwell's demon. In a letter to Peter Tait in 1867, James Clerk Maxwell presented a thought experiment that illustrates the *statistical* nature of the (then brand new) Second Law of thermodynamics. Maxwell invited us to imagine "two vessels divided by a diaphragm [that] contain elastic molecules in a state of agitation which strike each other and the sides", where one vessel contains more "agitated" molecules than the other, i.e. the gas on one side of the diaphragm is *hotter* than the other (Knott, 1911, p.214). Maxwell then asks us to imagine a hypothetical entity — later dubbed a *demon* by (a very enthusiastic) Lord Kelvin (1874) — who knows the properties ("paths and velocities") of each and every molecule in the two vessels and who can open the diaphragm and let single molecules through *without performing work*. This hypothetical entity can sort the molecules in the two vessels in any way it desires, so too in ways

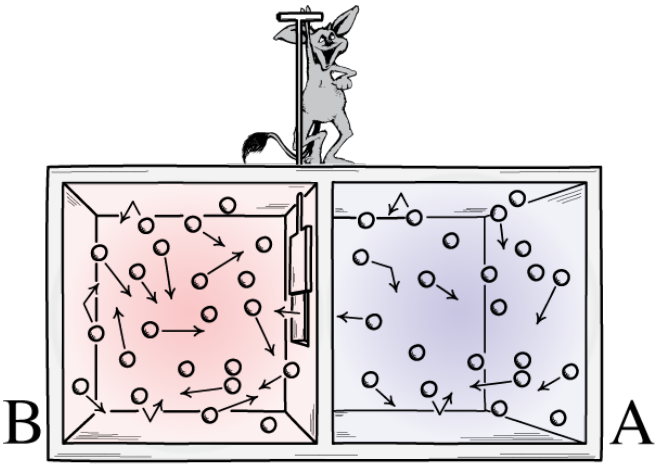

Figure 4.4: Maxwell's demon. (Image courtesy of John Norton; obtained from `sites.pitt.edu/~jdnorton/jdnorton.html`.)

that violate the Second Law of thermodynamics (e.g. by cooling one vessel down without performing work, or by using the sorting process to generate "motion of large masses" (*ibid.*), again, without performing work.) Thus Maxwell illustrated that the Second Law of thermodynamics *can* be violated and, hence, is *statistical*: it holds *typically*, but not *necessarily*.

I mention this STE primarily because few thought experiments have such a 'life of their own' as Maxwell's demon does. Maxwell conceived his STE to illustrate a violation of the Second Law of thermodynamics purely in the context of 'classical' statistical mechanics (Hemmo and Shenker, 2012). The direct responses and the attempts to 'exorcise' (i.e. prove the physical impossibility of) Maxwell's Demon from the early 1900s, notably by the physicist Marian Smoluchowski, were in the same spirit; see e.g. Ehrenberg (1967); Bub (2001); Rex (2017).

But later analyses of Maxwell's demon and the attempts to exorcise it soon took place in the context of entirely *different* scientific theories, notably in the context of (quantum) *information theory*; c.f. Leff and Rex (1990); Earman and Norton (1998, 1999); Bub (2001); Myrvold (2011); Norton (2013). The view on Maxwell's demon from the perspective of

information theory has been, and still is, widely influential (despite fervent attempts to prevent this by the above-cited authors), which is something that could never have been anticipated by Maxwell himself. This shows that, just like 'real' experiments, thought experiments too can have a "life of their own"; c.f. Hacking (1992).

(3) Einstein's photon-box. The debates between Einstein and Bohr at the heyday of quantum mechanics are legendary. Few interactions between scientists gave rise to so many famous thought experiments as the debates between Einstein and Bohr did. This, by itself, is noteworthy: at points in the history of scientific development where there is great *conceptual* discombobulation — that is, during *scientific revolutions* (Kuhn, 1977) — thought experiments come to the fore. One controversial thought experiment stands out in particular in the debates between Einstein and Bohr: Einstein's photon-box.

At the 1930 Solvay Conference, Einstein tried to disprove Heisenberg's uncertainty principle (that two non-commuting variables cannot both be determined with arbitrary precision) by means of a thought experiment: the photon-box. This thought experiment ran as follows. Imagine a box containing photons that is suspended from a spring; see Figure 4.5. Suppose that the weight of the box can be determined with arbitrary precision by reading the vertical position of the box. At some pre-determined exact moment in time, the box opens a shutter that releases exactly one photon. If we weigh (with sufficient precision) the box before and after the moment of release, then we can determine the energy of the photon to arbitrary precision. Given that we also know the exact moment of emission of the photon, we have now determined the value of two non-commuting variables with arbitrary precision: energy and time.

Bohr's reply to Einstein's photon-box, which was allegedly constructed during a sleepless night after Einstein presented his STE, appealed to general relativistic principles (which was *Einstein*'s own theory) and greatly shocked Einstein, but it did not convince Einstein completely — nor did it convince later commentators completely (Beller, 1999; de la Torre et al.,

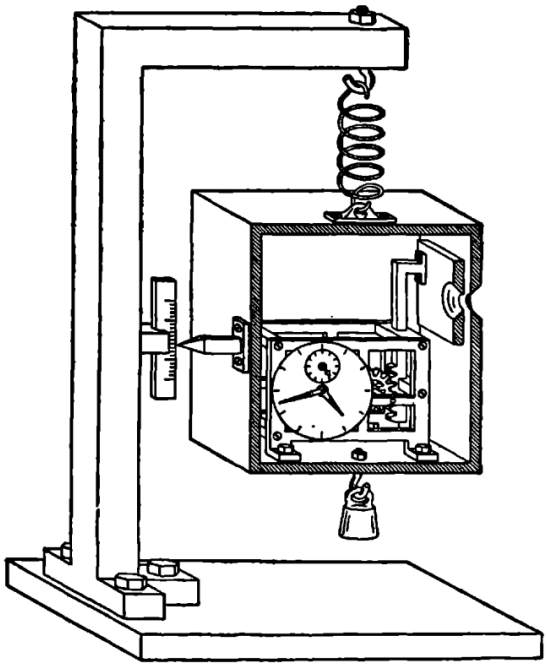

FIG. 8

Figure 4.5: Einstein's photon box. Image courtesy of Niels Bohr; obtained from (Schilpp, 1959, p.227).

1999; Marage and Wallenborn, 1999; Hilgevoord, 2002; Howard, 2007; Schmidt, 2022). The technical details of this STE are beyond the scope of this Chapter. What is important to note about the photon-box is that this is an STE with *conflicting interpretations*: two scientists are talking about the *same* STE, i.e. about the *same* imaginary scenario, while applying *different* scientific theories to this scenario.

(4) Norton's dome. Norton (2003, 2008) concocted a thought experiment that demonstrates an "unexpectedly simple failure of determinism in Newtonian mechanics". Norton asks us to imagine a point-mass at rest on top of a dome of a particular shape defined by the equation $h = (2/3g)r^{3/2}$, where h is the height of the dome, r is the radial distance from the center, and g is the gravitational constant. There is a standard solution in

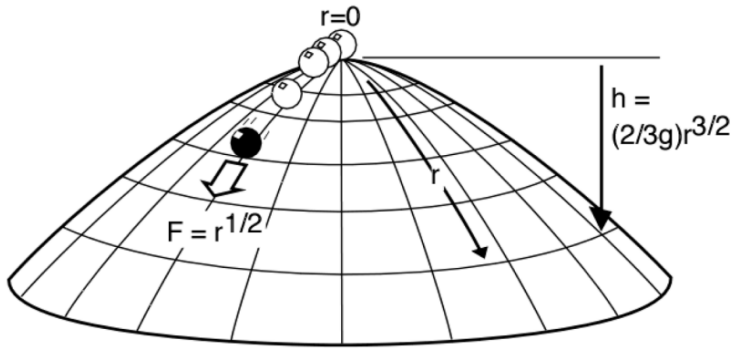

Figure 4.6: Norton's dome. Image obtained from (Norton, 2008).

Newtonian mechanics for the motion of the point-mass: the mass stays on top of the dome at rest, forever. This is no surprise: because there are no forces acting on the point-mass, we should expect that it stays at rest on top of the dome for all time. However, Norton shows that, due to the particular shape of the dome, Newtonian mechanics allows for *another* solution: a solution where the point-mass *spontaneously* starts moving down the dome at some random, arbitrary time. Thus, by demonstrating that this dome-scenario manifests an indeterministic scenario, Norton demonstrated that Newtonian mechanics is indeterministic.

Norton's dome has received a wide range of replies; see e.g. Malament (2008); Fletcher (2012); Laraudogoitia (2013); Van Strien (2014). Many replies attempted to 'disprove' Norton's dome by showing that the thought experiment goes wrong *somewhere*; see replies in (Norton, 2008). In any case, Norton's dome forced us to reconsider carefully what we mean when we say that a scientific theory is 'deterministic'.

I introduced Norton's dome because it illustrates two important features of STEs: (i) it introduces a scenario that *cannot* be realistic (because it involves point-masses, infinite times, etc.), with the sole aim of demonstrating a surprising feature of a *scientific theory*; and (ii) this STE does not concern the natural world, neither in design (the imaginary scenario of a ball waiting infinitely long on top of a dome until, at some arbi-

trary time, it spontaneously moves) nor with respect to its epistemic aim (demonstrating a surprising failure of 'determinism' in Newtonian classical mechanical theory). So, we have here a case of a thought experiment in natural science that, for all practical purposes, *does not concern the natural world at all*, neither directly nor indirectly.

4.2.3 Core question (I): what STEs are

Having presented the example STEs in the previous section, I now return to *core question* (I): what *are* scientific thought experiments? Given the familiarity that most of us will have with the term "thought experiment", this question is deceptively simple. As was evident from the lengthy list of scientific activities that STEs are closely related to (Section 4.2.1), the concept of scientific thought experiment refers to a wide, heterogenous range of scientific activities. To avoid ambiguity in my discussion, I next specify more precisely what I mean with the concept.

I take as my point of departure a provisional characterization of STEs that is based on the analysis of STEs by Gendler (2004):[93]

> To perform a scientific thought experiment is to reason about an imaginary scenario with an epistemic aim about the natural world. (4.1)

This characterization (4.1) rightly captures core characteristics of STEs, in that the performance of STEs always involves some type of reasoning about some type of imaginary scenario with some type of epistemic aim. That much is true. But I disagree with (4.1) in three respects, as I shall explain next.

Firstly, not all STEs have an epistemic aim pertaining to the natural

[93] Gendler (2004) herself characterises STEs as: "to perform a thought experiment in science is to reason about an imaginary scenario with the aim of confirming or disconfirming a hypothesis about the natural world." I chose to adjust Gendler's characterization because the epistemic aim of STEs can be something else than confirming or disconfirming a hypothesis (recall the lengthy list of aims from Section 4.2.1).

world. Some STEs merely aim to demonstrate a feature of our scientific theories or models without aiming at extrapolating this result into some token of insight about the natural world. Norton's dome was a case in point. I already discussed this point in Section 4.2.1: I provisionally limit my attention to thought experiments with epistemic aims *relevant for science* — which may or may not concern the natural world.

Secondly, Gendler mentions that, for her, "the fundamental notion [is] the *performance of a thought experiment*, with the notion of *being a thought experiment* derivative therefrom" (Gendler, 2004, p.1155). I take things to be the other way around: I hold that a thought experiment can be meaningfully said to 'exist' independently of whether or not someone performs it, in much the same way as a mathematical proof or an argument 'exists' even when no-one is currently reading or reasoning through them, and very much in the same way as a theoretical model 'exists' even when no-one is currently reasoning about it. A relatively uncontroversial way in which I shall say that STEs (and theoretical models) 'exist' is *in virtue of their description* — I will elaborate on this point below. This indicates an important difference between thought experiments and acts of imagination: thought experiments 'exist' whether or not someone performs them, while acts of imagination — defined as a sequence of mental states of imagination in Chapter 3 — exist only in the heads of individual imaginers. This brings me to my third point.

Thirdly, Gendler's characterization of STEs casts its net too wide. This characterization labels as STEs not only 'obvious' thought experiments from the history of science such as Galilei's falling bodies, Newton's bucket and Maxwell's Demon, but it also potentially labels as STEs also as *ad hoc, impromptu* imagination-based reasoning processes that we perform in daily life about questions with immediate practical relevance like "what would happen if I were to knock over this glass of water?" or "would I enjoy living in this house?" (Williamson, 2016) or "does this couch fit through that door?" (Dorsch, 2016b); recall the previous Chapter of this Thesis. These latter *ad hoc* imagination-based reasoning processes

are interesting in their own right, but I believe that we should not call them *thought experiments*. Not *every* epistemic act of imagination is a thought experiment. If it were, then the domain of philosophy of thought experiments and the domain of epistemology of imagination *in general* would be one and the same. But they are not the same.

I take it that STEs differ from acts of imagination in at least two ways: (1) thought experiments always have *descriptions* because they are meant to be *communicated*, whereas acts of imagination do not always have descriptions, and (2) thought experiments have inter-subjectively stable imaginary scenarios associated with them, which cannot be identified with a sequence of mental states of imagination, whereas acts of imagination *are* sequences of mental states of imagination (recall Chapter 3, Section 3.2, p.106). I elaborate on each in turn.[94]

(1) STEs have descriptions. This point may seem insubstantial but it really is not. I shall argue throughout this Chapter that the descriptions of STEs contribute crucially to our performance of STEs, epistemologically speaking. I indicate two ways in which they do so.

First and foremost, the descriptions of STEs are *deliberately designed instructions* for setting up a specific imaginary scenario in the mind. (The analogy with works of fiction bangs loudly on the door. I shall let it in soon.) As such, the descriptions of STEs *guide* our imagination in a specific direction — in a direction that helps us achieve the epistemic aim of the thought experiment (if it is a *good* description). In other words,

[94] I mention here that one may be tempted to construe thought experiments as *imagined scientific experiments* (if only because of the *name* "thought *experiments*"); see e.g. Sorensen (1992); Sartori (2023); c.f. De Mey (2003), who argues for the 'dual-nature view' that STEs are always experiments *and* arguments. In this light, to perform a thought experiment is to *imagine performing a scientific thought experiment*. We should then be able to increase our understanding of *thought*-experimenting by combining (i) an account of what it is to *imagine* (Chapters 2–3), with (ii) an account of what it is to perform a scientific experiment. While I agree that this strategy is helpful to some extent, I submit that it is too limited. Not all *thought experiments* are *imagined scientific experiments*. Maxwell's demon and Norton's dome are examples in point: it is hard to see how an account of scientific experimentation could increase our understanding of these thought 'experiments'. The reasoning that occurs when we perform thought experiments can be *different* than imagining performing an experiment. See more on this below.

the descriptions of STEs are descriptions of an imaginary scenario *and* explicit instructions for reasoning about that imaginary scenario with some given epistemic purpose, often in a (roughly) pre-determined way that should achieve this epistemic purpose effectively. STEs always have an identifiable "job to fulfill" (Hacking, 1992): they are supposed to teach us something specific and they were deliberately designed to fulfill this job in an effective way. This distinguishes STEs from other, non-thought-experimental epistemic uses of imagination.

Secondly, the descriptions of STEs make it possible to *coordinate our imaginings inter-subjectively*, in the sense that the *same* STE can be studied and performed by *different* people, even hundreds of years after the thought experiment was first conceived. As we saw with the responses to e.g. Newton's bucket and Einstein's photon-box, moreover, there is significant *variation* in the performance of an STE (Kujundzic, 1998). We regularly *agree* or *disagree* about thought experiments, both about the *content* of the thought experiment's imaginary scenario, and even about the *epistemic aim* that the thought experiment is supposed to help us achieve. As Mach's reply to Newton's bucket showed us, there can be ambiguity, and even disagreement, about the content of an STE's imaginary scenario. Moreover, in an important sense, we can *discover* things about the imaginary scenario of a thought experiment, like e.g. Bohr's reply to Einstein's photon-box showed us: we can *discover* which laws of nature govern the world of an STE, in a way that is much less arbitrary than voluntary choice. In other words: the imaginary scenarios of STEs are often *under-determined* by their description, but they can be *investigated* and features of the scenario that were not directly described in the description of the STE can be *discovered*. (The analogy with works of fiction bangs on the door even louder than before.)

And yet, through all this variation and change in STEs, it often does make sense to say that we are still talking about the *same* STE.[95] The content of Maxwell's demon changed radically through the years, but we

[95] This raises the issue of *identity conditions* for STEs; see Section 4.5.5 for discussion.

still call it Maxwell's demon. Likewise, Mach greatly disagreed with Newton about the *content* of his bucket STE, but they were discussing the *same* STE nonetheless. This brings me to the second way in which STEs differ from *ad hoc* acts of imagination.

(2) *STEs have topics; they have inter-subjectively stable imaginary scenarios associated with them.* I have just argued that STEs have descriptions and that these descriptions play important epistemic roles in our performance of thought experiments, in the sense that they are deliberately designed instructions for setting up a specific imaginary scenario, and that they make it possible to coordinate our imaginings inter-subjectively. I have also argued we can *communicate* inter-subjectively about the imaginary scenario that is associated with a thought experiment, and that we can *agree* and *disagree* about its content, and that we can *discover* things about this content that were under-determined by the STE's description.

Thus, I submit, the imaginary scenarios of STEs are *inter-subjectively stable* to some extent. This implies that these imaginary scenarios cannot be identified with *mental content*, i.e. with the content of an act of imagination of some particular performer of the STE; that would leave unexplained too many of the above-mentioned characteristics of STEs, notably the inter-subjective stability of the imaginary scenario of STEs and the concomitant fact that several people can be said to perform the *same* STE despite there being significant *variation* in the content of their respective acts of imagination.

Fortunately, because I assumed that STEs have descriptions, there is a natural way available of describing this inter-subjectively stable scenario of STEs. Descriptions of STEs are composed of *sentences*.[96] Sentences express propositions. Propositions have topics. The *topic* of the propositions expressed in the description of an STE *is* the inter-subjectively stable imaginary scenario of the STE.

[96] Descriptions of STEs are also composed of *pictures* and other artifacts with referential content (rather than semantic content). These pictures also *refer* to the STE's topic, providing referential content about it rather than semantic content. (See more below.)

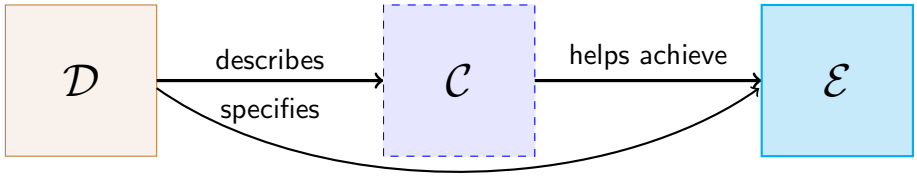

Figure 4.7: Standard schema for STEs: a description $\mathcal{D}$ describes a topic, which is an imaginary scenario $\mathcal{C}$, and specifies an epistemic aim $\mathcal{E}$ that should be achieved by reasoning about $\mathcal{C}$.

Adjusting the provisional characterization of STEs discussed above, I propose the following improved characterization of STEs:

Characterization of STEs:
To perform an STE is to reason, upon engaging with an STE's description, about the topic (an imaginary scenario) of that STE, with an epistemic aim that is relevant for science. (4.2)

On the basis of this characterization (4.2), I now distinguish *three* distinct 'ontic components' of STEs:

(i) The *description* of an STE,

(ii) The *topic* of that STE,

(iii) The *epistemic aim* of that STE.

With these three 'ontic components' of STEs in hand, I am now in a position to *define* what an STE is. For the sake of convenience, I shall employ the language of set-theory to do so, as follows. A *scientific thought experiment* STE is an ordered triple

$$\mathsf{STE} = \langle \mathcal{D}, \mathcal{C}, \mathcal{E} \rangle, \tag{4.3}$$

consisting of a *description* $\mathcal{D}$, its *topic*, i.e. the *imaginary scenario* $\mathcal{C}$ described by $\mathcal{D}$, and an *epistemic aim* $\mathcal{E}$. See Figure 4.7 for an illustration.

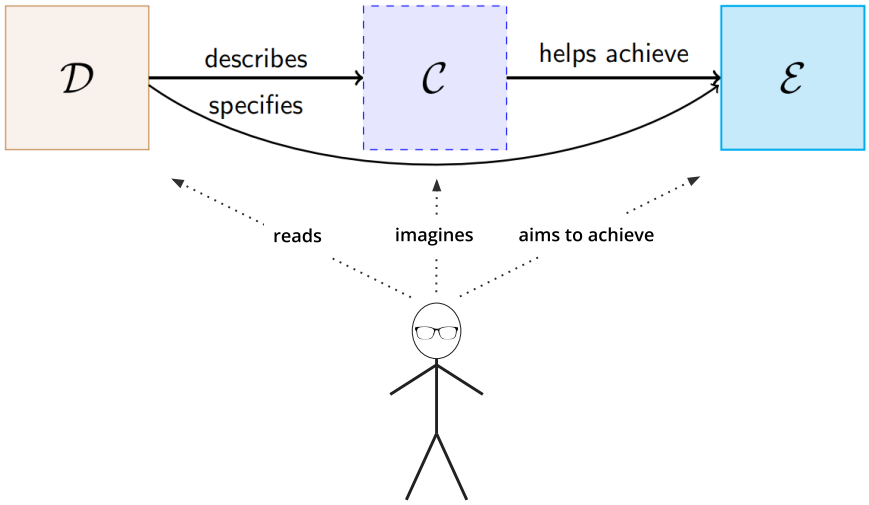

Figure 4.8: Performing an STE.

Using definition (4.3) and my proposed characterization of STEs (4.2), I next explicate what it means to *perform* an STE:

> [Performing an STE] Subject S *performs* STE $= \langle \mathcal{D}, \mathcal{C}, \mathcal{E} \rangle$ iff upon engaging with description $\mathcal{D}$, S reasons about imaginary scenario $\mathcal{C}$ (i.e. the topic of $\mathcal{D}$), with epistemic aim $\mathcal{E}$. (4.4)

See Figure 4.8 for an illustration.

I next elaborate on these three components $\mathcal{D}$, $\mathcal{C}$ and $\mathcal{E}$ and I indicate what an account of STEs should explain about them.

(i) The description of STEs. For the sake of simplicitly, I shall assume that the description of an STE is a *concrete object* that typically consists of several lines of text on paper or screen (with semantic content), and which is often supplemented by equations and audiovisual media such as images, pictures, diagrams, videos and what have you (with representational content). The performer of the STE *reads* the description — or it observes it, feels it, or engages with it in any other way such that the performer

understands the imaginary scenario that the description describes.

Descriptions of STEs describe the *topic* (the imaginary scenario) of an STE. But, as should be evident from the examples discussed in the previous Section, descriptions are often only *partial* descriptions of their topic; descriptions of STEs *under-determine* their imaginary scenario. Often descriptions of STEs do not mention crucial assumptions that are part of the imaginary scenario of the STE. The assumption in Newton's bucket that the bucket is spinning in otherwise-empty space was a case in point. An account of STEs should explain how this works; i.e. an account of STEs should explain how the description of an STE 'gives rise' to its (under-determined) inter-subjectively stable imaginary scenario. Additionally, I submit, an account of STEs should explain what, if anything, the description of an STE contributes to a subject's mental process of performing an STE (4.4) and achieving its epistemic aim.

(ii) The imaginary scenarios of STEs. The topic of an STE is an imaginary scenario. Imaginary scenarios are, well, *imaginary*. This, by itself, is somewhat of a problem: it remains highly controversial what an imaginary scenario is, how they can be inter-subjectively stable, and how we engage with them, even in light of Chapters 2 and 3 of this Thesis. One notable problem here is how to determine the *content* of the imaginary scenario of an STE: how do we determine the (under-determined) features of the *topic* of an STE?

The imaginary scenarios of STEs give rise to many epistemic puzzles. The imaginary scenarios of STEs often contain many *idealizations* (point-masses, frictionless slopes, perfect spheres, elastic collisions, etc), they are importantly *incomplete*, in the sense that some things are the case in the imaginary scenarios of TEs but many other things are undetermined (e.g. there exist no facts of the matter about the star Betelgeuze in the imaginary scenario of Maxwell's demon), and they often contain *physically impossible* entities, such as rotating buckets in otherwise empty space and demons that know the exact properties of all molecules in a gas. As such, the imaginary scenarios of STEs are often flat-out *physically*

impossible: they are scenarios that cannot exist in the natural world. An account of STEs should explain, then, how it is possible that we can learn about the *natural world* by engaging with such idealised, incomplete and often *impossible* imaginary scenarios. I shall argue in Section 4.3 that the language of fiction enables us to explain this straightforwardly.

In my explication of what it means to perform an STE (4.4), I wrote that "S *reasons* about imaginary scenario $\mathcal{C}$". What does it mean to reason about an imaginary scenario? I have already noted (fn.94, p.217) that the reasoning present in the performance of STEs often encompasses more, or just something else, than "imagining performing a scientific experiment". An account of STEs should explain what this reasoning *does* encompass.

I mention here one notable aspect of the reasoning that occurs when we perform an STE, which relates to the under-determination of the STEs imaginary scenario $\mathcal{C}$ by the STE description $\mathcal{D}$, discussed above. The reasoning that occurs when we perform an STE should of course be *based* on the premises expressed in the STE description; it should not be *contrary* to the premises expressed in the description — to do so would be to *refuse* to perform the thought experiment. But, just like the imaginary scenario of an STE is often under-determined by its description, so is the reasoning process that should occur when we perform the STE under-determined by its description. Galilei's falling bodies was a case in point, as, to reach Galilei's intended conclusion, some additional reasoning beyond what is prescribed in the description is required (see also Section 4.5.2 for elaboration). Clement's Sisyphus certainly was a case in point, as it left the reasoning process *entirely* open. To give one clear example: the conclusion from Norton's dome should be reachable whether we imagine the dome colored or colorless, and whether we imagine the ball rolling north or south or east or west, and so on, but we *must* imagine the shape of the sphere exactly as it is described, else the conclusion does not follow.

This raises the question which forms of reasoning are valid ways of reaching the epistemic aim of any given STE. It seems that, at least for some STEs such as Clement's Sisyphus, the epistemic aim of the STE

should be *reachable* under significant variation in the under-determined features of the imaginary scenario of an STE *and* of under significant variation in the under-determined features of the reasoning process prescribed in the description of an STE. An account of STEs should be able to explain how we seem to achieve the epistemic aims of STEs so efficiently *despite* — or, perhaps, *because of* — this invariance. (I shall argue that the *invariant* aspects of STEs are facts about *scientific models*.)

(iii) The epistemic aim of STEs. The epistemic aim of a given STE is some specific token of insight relevant for science — a token of knowledge or understanding about a scientific theory, model or concept, or about some real-world phenomenon or class of phenomena. On the basis of the explanations for components (i) and (ii) just mentioned, an account of STEs should explain *which* types of epistemic aims STEs could in principle give us access to, and it should explain *how* we can reach these epistemic aims by specifying which mental processes occur when we perform STEs and which epistemic standards they must meet. This brings me to the second core question concerning STEs: what, and how, we learn by performing STEs.

4.2.4 Core question (II): what, and how, we learn by performing STEs

The second core question concerning STEs was: what, and how, do we learn by performing STEs. Using the definition of STEs (4.3) and the explication of the performance of STEs, this question assumes the following form: which types of epistemic aims $\mathcal{E}$ can be achieved by performing an STE, and how do the description $\mathcal{D}$ and the imaginary scenario $\mathcal{C}$ of the STE help us in achieving these aims?

In Section 4.2.1, I presented a list of epistemic aims that arguably can be achieved by performing STEs. I make two comments about this list.

The first thing to note about this list is that it includes so many different epistemic aims that it is not practically achievable for an account of STEs to explain in detail how we reach each and every one of these

aims individually. To strive for this would be the wrong strategy to take. It is however also clear from this lengthy list that focusing on explaining how we can achieve just one or several of these epistemic aims by performing STEs would leave unexplained many others. This too is the wrong strategy to take. I suggest that the best strategy for formulating a comprehensive account of STEs is to look for scientific activities, closely related to thought-experimenting, that achieve a wide variety of similar epistemic aims, and to connect thought-experimenting to *that* activity. If we can plausibly argue that the practice of thought-experimenting falls under that other activity, then we already made big steps towards explaining how *STEs* can achieve these aims too. In this Chapter, specifically Sections 4.4 and 4.5, I argue that *model-based reasoning* does the trick.

The second thing to note is that an answer to the question *which* epistemic aims we can achieve by performing STEs is only half of the story. The other half of the story concerns *how* we achieve those aims. Two concepts are important here: (i) it is often argued that STEs have more *heuristic value* than non-thought-experimental forms of reasoning, in the sense that we can gain insight *easier* and more *effective* by performing STEs than through other means, such as 'mere' non-imaginative reasoning through an argument; and (ii) that STEs can have a distinct type of *demonstrative force*, which makes the conclusions drawn from performing STEs particularly convincing or perhaps even uniquely *justified* (Brown, 1991; Arthur, 1999; Glas, 1999; Gendler, 1998, 2004; De Mey, 2006; Camilleri, 2014a, 2015). An account of STEs must explain the source of both the heuristic value and demonstrative force of STEs.

De Mey (2006, p.227, fn.10) notes that the heuristic value and the demonstrative force of STEs are two sides of the same coin, in that "[o]ne might very well argue that an account of the heuristic value of thought experiments will also explain why they have "demonstrative force," or vice versa". But De Mey goes on to describe how "in practice, however, these are different debates. Naturalistic philosophers like Nancy J. Nersessian deploy findings from cognitive science to account for the heuristic force

of thought experiments. Analytic philosophers like Tamar Gendler focus more on demonstrative force. I believe an integration of such approaches is not only possible but also desirable" (De Mey, 2006, p.227, fn.10). I believe so too: in Section 4.5 I shall use my account to integrate these two approaches and explain the heuristic value and demonstrative force of STEs from a single perspective.

The paradox of thought experiments

At this point, I should mention one particular topic of debate that dominated the analysis of STEs for decades: the *paradox of thought experiments* (Horowitz and Massey, 1991); c.f. Stuart (2015). In the introduction of his influential 1964 paper on STEs, Kuhn posed a question that is now central to the paradox of thought experiments (Kuhn, 1977, p.241):

> How, relying exclusively upon familiar data, can a thought experiment lead to new knowledge or understanding of the world?

This question by itself expresses no obvious paradox.[97] Indeed, I have argued in the previous Chapter of this Thesis that imagination is a source of knowledge of the natural world in at least *four* distinct ways; recall Chapter 3, Sections 3.5.2 (p.188) and 3.5.3 (p.190). Yet when thought experiments became the subject of systematic study in philosophy of science in the 1990s, almost immediately a significant part of the literature turned into a debate which eerily resembled the eternal 'empiricism-versus-rationalism' debate in epistemology: Brown's rationalist account of STEs versus Norton's empiricist account of STEs.

Brown (1991) introduced a taxonomy of thought experiments with one special category: Platonic thought experiments. Platonic thought experiments simultaneously provide a destructive argument against an existing theory and suggest constructively a new theory — Galilei's falling

[97] Moreover, Kuhn himself provided an answer to this question in accordance with his familiar view on scientific revolutions: by performing a thought experiment, a scientist can be made to realise that his concepts are inadequate for describing situations he has already confronted, and so "the scientist learns about the world as well as about his concepts." (Kuhn, 1977, p.261)

bodies being the paradigm example (Brown, 1991, p.41). According to Brown, new knowledge resulting from Platonic thought experiments such as Galilei's falling bodies is synthetic *a priori* knowledge of laws of nature, which, in conjunction with his Platonic view on laws of nature, can only be accounted for by granting thought experiments access to the Platonic realm of universals: by performing thought experiments, Brown argued, we can gain unique access to Platonic truths. Norton shunned this "epistemic magic" and defended the opposing, hard-nosed empiricist account (e.g. Norton (1993, 1996, 2004a,b)): STEs are nothing but *arguments*, we can learn from them nothing more than what can be legitimately inferred from what we already know.

In his description of the history of the paradox of thought experiments, Stuart writes (Stuart, 2015, p.6):

> The transition from puzzle to paradox takes place when Kuhn's open-ended question transforms into a dilemma between two options: a world with epistemic magic, and one without. [...] It becomes a question with conflicting but well-credentialed answers. Given that thought experiments provide or purport to provide information about the physical world, yet do not require new information about the physical world, either the new information is a rearrangement of old data, or else it comes from rational insight.

Both Norton's and Brown's accounts have been scrutinised in the literature and currently neither are widely accepted in the form in which they were originally proposed. Yet, as Stuart reports, it is a remarkable fact that between 1991 and 2009 at least 69% (!) of the relevant literature mentions a paradox in the abstract or introduction (Stuart, 2015, p.8). Interestingly, however, the focus on 'paradox' declined from 2009 onward. Stuart purports — quite plausibly — that this is due to the increase cognitive psychology-based accounts of STEs. Indeed, most contemporary authors do not consider there to be a deep epistemological puzzle pertaining to the idea that STEs do not involve novel empirical data yet we seem to learn about the natural world by performing them. These days, most

authors agree that STEs are just "a species of reasoning rooted in the ability to imagine, anticipate, visualise, and re-experience from memory" (Nersessian, 2018, p.310). This is my view of STEs too, as should be evident from my analysis of the Question of Knowledge Through Imagination from Chapter 3, particularly Section 3.4, and as should become evident from the account of STEs that I propose in this Chapter.

I wish to emphasise here that *some*, but *not all*, STEs have epistemic aims that concern insight about the natural world — their 'target system' is the natural world, as they say. The 'paradox of thought experiments' pertains to these STEs only. Other STEs, such as Norton's dome, do *not* have epistemic aims that concern the natural world; hence, these STEs do not suffer from the 'paradox', even if it were a genuine paradox (which it is not). This fact was often overshadowed by the remarkable focus on "paradox" in the late-1990s literature on STEs, but fortunately it has become more obvious in recent years that not all STEs are *prima facie* paradoxical.

The difference between STE-beliefs and quasi-perceptual beliefs

Additionally, I wish to note that beliefs gained by performing STEs cannot and should not simply be regarded *quasi-perceptual beliefs* (3.9), which were the topic of Chapter 3. It is true that quasi-perception is often relevant in the performance of STEs. But there is *more* relevant in the performance of STEs rather than just quasi-perception. I submit that there are at least three important differences between beliefs gained by performing STEs — call them STE-beliefs — and quasi-perceptual beliefs.

Firstly, there is a difference in *topic* of the respective beliefs. In analogy with ordinary perceptual beliefs, quasi-perceptual beliefs concern beliefs about relatively simple, theory-neutral, *perceivable* matters of fact, e.g. beliefs about the amount of elephants that fit in your room or about the amount of windows in the house you grew up in; recall the Examples in Chapter 3, Section 3.3.3. STE-beliefs, by contrast, are *scientific* beliefs. As the examples in Section 4.2.2 demonstrated, STE-beliefs are typically

not beliefs about ordinary perceivable matters of fact. They are complex, theory-laden, *inferential* beliefs about *scientific* matters: e.g. beliefs about the speed of falling of objects of unequal weight, about the definition of relative motion or about the nature of the Second Law of thermodynamics. While some STE-beliefs do concern matters of fact about the natural world, they do so only *indirectly*: they are, first and foremost, *beliefs about scientific theories and models*. Moreover, STE-beliefs need not *concern* the natural world at all, neither directly nor indirectly, as Norton's dome showed us. Not all scientific theories and models represent the natural world — some models are just constructed to sharpen our mathematical tools, for example, or to explore the inter-dependence of scientific concepts. When we perform an STE about *these* scientific constructs, then, even if quasi-perception is present in the performance of the STE, the result will not be a belief about the natural world.

Secondly and relatedly, there is a difference in *content* of the respective imagined scenarios. The imagined scenarios that yielded quasi-perceptual beliefs considered in the previous Chapter did not only accurately represent the natural world, they *pictured* it; recall Chapter 2, Section 2.5.5. As such, the propositions that came to mind on the basis of this quasi-perceived scenario *were* propositions about the natural world. STE-beliefs are different. As I have argued in Section 4.2.3, the imaginary scenarios of STEs contain many idealizations, deliberate falsehoods, and non-existent and even *impossible* entities: *the imaginary scenarios of STEs often do not picture the natural world*. Thus, the propositions that come to mind upon performing an STE often do not do not concern the natural world at all; they must first be *transformed* into propositions that do concern the natural world. The representation-relation between the imaginary scenario of an STE and the natural world is of course crucial here, but it requires a more nuanced type of representation than mere picturing. I shall introduce an account of representation that can adequately deal with STEs in Section 4.3.4.

I note moreover that, whereas (3.9) involved only mental states of *ac-*

tion-imagination (2.31), we would do wise to allow for the performance of an STE to involve also mental states of *proposition*-imagination (2.13). The relevant mental states of imagination in thought-experimenting are not *only* mental states of imagination with *mental imagery*, i.e. states of action-imagination, they are also mental states of imagination with *semantic* content, i.e. states of proposition-imagination. The descriptions of STE explicitly *provide* us with propositions that should be imagined: e.g. to imagine *that* a photon can be released at some pre-determined time from a photon-box. Mental imagery may be present in such a mental state of imagination, but arguably the semantic content does most of the epistemic work in the performance of some STEs; c.f. Salis and Frigg (2020). To give an extreme example: consider performing a thought experiment where you imagine having *no* sensory modalities: you can imagine this propositionally, but how would you imagine this *imagistically*?

Thirdly, there is a difference in the *process* of obtaining the respective beliefs. I reconstructed the process of obtaining quasi-perceptual beliefs as a two-step process (3.9), consisting of a quasi-perceptual step and an inferential step about the accuracy of one's imaginings. The process of obtaining STE-beliefs is often much more complex. Sure, there is often quasi-perception at play in the formation of an STE-belief, which may even play a *crucial* role in the formation of the belief, as e.g. Clement's Sisyphus showed us. But typically there is are many *more* processes going on in the formation of STE-beliefs than in the formation of quasi-perceptual beliefs. I mention four distinct processes: (i) one engages with a description that specifies some thought-experimental imaginary scenario and imagines that scenario, (ii) one must imagine this scenario *whilst* having the right scientific theories and models in mind, i.e. one must think of the theories and models that are relevant for the imaginary scenario of STEs and for its epistemic aim; (iii) one needs to explicitly relate the results of their thought experiment to the 'content' of scientific theories and models before one can gain a belief about the latter, and this process is rather more involved — *and much more inferential* — than mere quasi-perceptual de-

termining of matters of fact; (iv) and *only then* can one use this belief to gain beliefs about the world.

I next review two accounts of STEs that currently are amongst the most popular accounts of STEs: the *argument view*, which is a family of accounts of STEs that holds Norton's empiricist spirit high, and the *mental-modeling view*, which is a family of accounts of STEs that employs the notion of "mental models" popularised by Johnson-Laird (1983). I discuss each in turn.

4.2.5 STEs between arguments and mental models

The argument view

A long-standing account of STEs insists that STEs are simply some type of argument or form of argumentative reasoning: call this *the argument view of STEs*. The primary motivation of the argument view is the empiricist stance that the only legitimate bases for novel insight into the natural world are *empirical observations* and *arguments*; hence, because STEs are performed in the imagination and therefore do not involve novel empirical observations, the only way in which STEs *could* provide novel insight is through presenting an argument. In terms of the two *core questions* of STEs, mentioned in the Introduction of this Chapter, the argument view holds that (I) STEs *are* arguments, and (II) we can learn about the world by performing an STE because to perform an STE is to reason through an argument, which is a valid way of learning about the natural world. Admittedly, the argument view is "tantalizingly elegant because it rests on the uncontroversial claim that reasoning through acceptable argument forms is a source of knowledge" (Meynell, 2014, p.4154).

The strongest version of the argument view, championed by Norton (1993, 1996, 2004a,b), holds that the role of STEs is just to convey arguments and, moreover, that *only* the arguments conveyed via STE have epistemic value. The fact that imagination is involved in performing STEs is epistemically irrelevant. The role of the imaginary scenario of an STE is

only to provide some rhetorical 'confetti', as it were, to the argument: according to Norton, STEs are mere "picturesque arguments and in no way remarkable epistemically" (Norton, 1996, p.334). Weaker versions of the argument view do not explicitly deny the psychological or epistemological role of imagination in STEs, but they nonetheless focus in their evaluation of STEs almost exclusively on the (validity of the) arguments that are presented via STEs; see e.g. Sorensen (1992); Häggqvist (2009); El Skaf and Imbert (2013); El Skaf (2018, 2021); El Skaf and Palacios (2022); Mulder and Muller (2023). In short: the argument view holds that we can learn from STEs nothing over and above what we can learn from arguments.

There are many problems with Norton's argument view, several of which we already came across in the previous Chapter (Section 3.4.4). These problems are instructive to re-review for the present purpose. To begin: it is regularly noted that the claim that "thought experiments are arguments" is ambiguous. Brendel (2018) has identified no less than *six* distinct claims within it (five 'main' theses, of which one divides into two sub-theses), of which some are much less plausible than others. I present these claims here as reconstructed by Brendel (2018, p.238), summarised conveniently where possible. In Section 4.5, I shall put forward analogous claims for my proposed account.

(1) **Identity Thesis.** STEs are type-identical with arguments.

(2) **Reconstruction Thesis.** STEs "can always be reconstructed as arguments based on explicit or tacit assumptions that yield the same outcome" (Norton, 2004a, p.1142).

  (2a) **Reliability Thesis.** "If thought experiments can be used reliably epistemically, then they must be arguments (construed very broadly) that justify their outcomes or are constructible as such arguments" (Norton, 2004b, p.52). A thought experiment is a "reliable mode of inquiry" only if the argument into which it can be reconstructed justifies its conclusion.

  (2b) **Elimination Thesis.** "Any conclusion reached by a (successful)

scientific thought experiment will also be demonstrable by a non-thought-experimental argument" (Gendler, 2010, p.34).

(3) **Epistemic Thesis.** STEs and the arguments associated with them have the same epistemic reach and epistemic significance. An STE epistemically justifies its outcome to the same degree as its associated argument justifies its conclusion.

(4) **Empirical Psychological Thesis.** To perform an STE is to reason through an argument.

(5) **Empiricist Thesis**: The result of a thought experiment can only come from experience: "The result of a thought experiment must be the re-formulation of [...] experience by a process that preserves truth or its probability." (Norton 2004a, 1142).

Of these Theses, only the Empiricist Thesis (5) is widely endorsed by most contemporary authors; it is endorsed even by those who reject Norton's argument view. The only account that seems to explicitly reject this thesis is Brown's (1991; 2004) aforementioned Platonic account of STEs, according to which some STEs provide us with privileged access to the Platonic world of universals. Brown's view does not have many supporters because nearly all contemporary authors prefer a *naturalistic* account of STEs: as I noted in the previous Section, these days, everybody seems to agree STEs are just "a species of reasoning rooted in the ability to imagine, anticipate, visualise, and re-experience from memory" (Nersessian, 2018, p.310). STEs do not provide us with *novel* 'input' that comes from a source external to the performer of the STE: the only 'input' for an STE is the content of an *STE description* and input that the performer of the STE brings to the STE herself, e.g. scientific knowledge, background beliefs, memories, social conventions, 'imaginative constraints' rooted in our sensory and motory processing mechanisms, etc., recall Chapter 3.

The Identity Thesis (1) and the Empirical Psychological Thesis (2) have met most resistance in the literature, for multiple reasons. Firstly, El Skaf and Stuart (2023) note that, to say that 'STEs *are* arguments'

makes sense epistemologically, but that it often remains unclear in these discussion what exactly an *argument* is, ontologically speaking. (With respect to my definition of STEs (4.3), this objection becomes even more poignant.) More importantly, it is undeniable that the *performance* of STEs (4.4) is at least *psychologically* distinct from mental processes that amount to 'mere' reasoning through an argument. The performance of STEs is an act of imagination that often crucially involves *mental imagery*, but a mental process of reasoning through an argument does not crucially involve mental imagery. Hence, the Empirical Psychological Thesis (4) is false. And if two mental processes are psychologically distinct, then they are not type-identical. Hence, the Identity Thesis (1) is also false.

The Reconstruction Thesis (2) and the Epistemic Thesis (3) are plausible on first sight, but they too are false. Admittedly, it is a good *methodological* rule that, in order to evaluate an STE, we should always begin by reconstructing the arguments that underlie the STE because a "precise argumentative reconstruction of a thought experiment can reveal merits and shortcomings of a thought experiment" (Brendel, 2018, p.291). This is what the proponents of 'weaker' argument views argue for, and which seems to be supported by many contemporary authors — if only to reveal where an STE goes *beyond* an argument. The account of STEs that I propose in this Chapter will hold this spirit high. However, in the previous Chapter, Section 3.4, I already discussed why the Reconstruction Thesis (2) is often rejected in the literature: if Miyazono and Tooming's (2023a) INACCESSIBILITY Claim is true — i.e. if we can obtain *justified* beliefs via an act of imagination (e.g. an STE) that we *cannot* obtain otherwise —, then the Reconstruction Thesis is false, and so it the Epistemic Thesis for that matter. I argued that the INACCESSIBILITY Claim is indeed true, and other authors did too. Recall that Gendler (1998, p. 415) explicitly rejected these Theses (2) and (3) when she wrote:[98]

> We have stores of unarticulated knowledge of the world which is not organized under any theoretical framework. Argument will not give

[98] See also Mach (1897, 1960); Kuhn (1977).

> us access to that knowledge, because the knowledge is not propositionally available. Framed properly, however, a thought experiment can tap into it, and — much like an ordinary experiment — allow us to make use of information about the world which was, in some sense, there all along, if only we had known how to systematize it into patterns of which we are able to make sense.

Moreover, particularly the Reliability Thesis (2a) and the Epistemic Thesis (3) conflict strongly with the ideas, which I described in Section 4.2.4, that STEs have more *heuristic value* than non-thought-experimental forms of reasoning, in the sense that we can gain insight *easier* and more *effective* by performing STEs than through other means, such as 'mere' non-imaginative reasoning through an argument; and they also conflict strongly with the idea that STEs can have a distinct type of *demonstrative force*, which makes the conclusions drawn from performing STEs particularly convincing or perhaps even uniquely *justified* (which relates to the INACCESSIBILITY Claim of Miyazono and Tooming (2023a))); c.f. (Brown, 1991; Arthur, 1999; Glas, 1999; Gendler, 1998, 2004; De Mey, 2006; Camilleri, 2014a, 2015).

Alongside these objections to the argument view, I wish to point out another issue with this account. The issue is that, even if it is true that 'STEs are arguments', then it is not immediately obvious what these arguments conveyed by STEs are *about*. The imaginary scenarios of STEs are akin to the wild worlds of fiction and fantasy: they include impossible objects and scenarios, non-existing entities, non-actual laws of nature, and so on and so forth; they are often *impossible* scenarios. This is a problem for the argument view, because we can only learn about the world via *sound* arguments, but sound arguments must be based on *true* premises about the world. But, if a thought-experimental scenario contains many deliberate *falsehoods*, how, then, could it 'be', or convey, a *sound* argument?

At the very least, I submit therefore, the proponent of the argument view must admit that there are *two* (mental) processes going on in the

performance of an STE: (i) one of the processes is an act of imagination, i.e. the imagined manipulation of a quasi-perceived scenario (recall Chapter 3, Sections 3.3–3.4), which may involve impossible scenarios, non-existing entities and non-actual laws of nature and the like; *and* (ii) then there is a process of reasoning, perhaps a process of explicitly reasoning through an argument, about how these *imagined* things relate to the natural world and what we can learn from this relation. This harks back to the two-step process for quasi-perceptual beliefs (3.9) that I elaborated on in the previous Chapter: STEs crucially involve *both* a *quasi-perceptual* process *and* an *inferential* process. Because the argument view ignores the first step of this two-step process, it does not explain *how* STEs present the sound arguments that teach us about the natural world.[99]

All things considered, while the argument view is 'tantalizingly elegant' in spirit, many of the Theses that make up this account are untenable. We need an account of STEs that does justice to the psychological and epistemic role of imagination in STEs, i.e. an account that can explain the heuristic value and demonstrative force of STEs in a naturalistic way. By far "the most promising" (Brown and Fehige, 2019, §5) and well-developed naturalistic family of accounts of STEs is the mental-modeling view of STEs. To this view I turn next.

The mental-modeling view

The mental-modeling view of STEs was originally proposed by McMullin (1985); Miščević (1992, 2007); Nersessian (1993, 1999, 2018) and is currently at least partially endorsed by most philosophers thinking about STEs, notably e.g. Gendler (1998, 2004); Cooper (2005); Camilleri (2014b); Clement (2009a,b, 2018); Brown and Fehige (2019).

A *mental model* is an imagined "structural analog of a real world or imaginary situation, event, or process [that] embodies a representation of the spatial and temporal relations among and the causal structure con-

[99] Sartori (2023) makes a similar objection to the argument view.

necting the events and entities depicted" (Nersessian, 1993, p.293).[100] According to the mental-modeling view of STEs, to perform an STE is to construct, manipulate and reason about a mental model with an epistemic aim. Importantly, it is argued that we can learn things through mental-modeling that we could not learn from mere argumentative reasoning because the manipulation of a mental model "affords epistemic access to certain features of current representations in a way that manipulating propositional representations using logical rules cannot" (Nersessian, 2018, pp.319-20); recall the quote from Gendler in the previous Section.

Proponents of the mental-modeling view of STEs often emphasise that mental models do not have (only) *propositional* content and, hence, that mental-modeling is *not* explicit, argumentative (i.e. propositional) reasoning. Instead, mental-modeling involves manipulating (non-propositional) structural *representations* of events and entities, which is typically just understood as *imagined* manipulation of *imagined* (and remembered) structural representations of events and entities, not as explicit processes of inferential reasoning. Thus the mental-modeling view goes beyond the argument view in acknowledging the psychological and epistemic role of imagination in thought-experimenting. This of course coheres well with our experience of *what it is like to perform an STE*: the subjective experience of performing STEs can often really be described as "quasi-observational" (Gendler, 2004), "experiential" (Dokic and Arcangeli, 2015) and even "embodied" (Gooding, 1992, 1993, 1994; Steier and Kersting, 2019; Kersting et al., 2021).

In terms of the two core questions concerning STEs (I) and (II), the mental-modeling view of STEs holds that (I) STEs *are* a type of reasoning grounded in our ability for mental-modeling, and (II) by performing STEs, we can learn about the world more than we can learn by mere reasoning through arguments, because by performing STEs we produce and justify novel beliefs about the natural world *quasi-perceptually*, in the sense that contemplating and manipulating imaginary scenarios evokes

[100] The term "mental model" was popularised by the work of Johnson-Laird (1983).

"quasi-sensory intuitions that could lead us to form new beliefs via [the] 'quasi-observational' imagistic kind of reasoning" that is characteristic of mental-modeling (Gendler, 2004, p.1161).

The mental-modeling view takes great steps in explaining the heuristic value and demonstrative force of STEs. When we construct and manipulate mental models, we make heavy use of our "perceptual and motor systems" (Nersessian, 2018, p.317), which are the neural mechanisms responsible for predicting and processing perceptual and motory sensory experience. As I have discussed many times throughout this Thesis, it is a remarkable fact that these mechanisms can be employed *in the imagination* in a functionally similar way as in ordinary perception — even to the extent that they can perform *epistemically* similar roles. The manipulations that we perform on mental models often 'mirror' the manipulations that we perform on *real* objects: our mental-modeling is heavily *constrained* by our 'innate motor schemas' and our past experiences with manipulating objects in the real world.[101] Quite plausibly, then, these mechanisms are responsible for the forceful and convincing sense that what happens in our mental models occurs *just like* it would occur in the real world: it just 'feels right'. This explains at least partly the *heuristic value* and *demonstrative force* of STEs — and it explains it in a way that the argument view cannot, I add.

This also pertains directly to the second step of the two-step process for quasi-perceptual beliefs (3.9) that I described in the previous Chapter, i.e. the step where we make inferences on the basis of our meta-beliefs about the *accuracy* of our imaginings. Indeed, the constraints on our imagination that are directly provided by mental models (the 'innate motor schemas', etc.) are argued to be *proper constraints* (recall Section 3.4.2), in the sense that they positively contribute to our imaginings accurately representing the natural world, thus enabling us to *learn* about the natural world by manipulating mental models. If these (proper) constraints provided by mental models are not propositionally available to a subject who performs

[101] See especially (Nersessian, 2018) for many references to empirical research that supports this claim.

an STE, then that subject can learn things about the world by performing an STE that they *cannot* learn by reasoning through an argument. This sheds yet another light on the demonstrative force of STEs, which in this light can be understood as *non-inferential justification* for the beliefs gained by performing the STE (i.e. justification in virtue of the quasi-perceptual process rather than in virtue of the inferential process).

But there is more. Until now, in this brief overview of the argument view and the mental-modeling view of STEs, I did not discuss the psychological and epistemic importance of *descriptions* of STEs. The argument view regards descriptions of STEs rather neutrally, dismissing them as 'mere' vehicles for conveying the semantic content of an argument. The mental-modeling view performs much better in this regard because it directly explains (at least partially) what and how the descriptions of STEs contribute to the heuristic value and demonstrative force of STEs, thus explaining (at least partially) their psychological and epistemic importance: the description of an STE is a "narrative [that] functions as a kind of user-manual for building the [mental] model" (Brown and Fehige, 2019, §4.6); c.f. (Matravers, 2014, Ch.5–6). A narrative plays a crucial role in *guiding our imagination*; and, if the imagination plays an important (and perhaps irreducible) psychological and epistemic role in gaining insight through thought-experimenting, then the descriptions of STEs *also* play a crucial psychological and epistemic role therein. This is not accounted for by Norton's argument view, where descriptions of STEs are regarded as a mere syntactic vehicle for conveying the semantic content of an argument; the narrative that specifies the (to-be) imagined scenario is regarded as nothing but epistemically irrelevant fluff. I like to see this improvement of the argument view by the mental-modeling view of STEs as a vindication of Feyerabend's (2020, p.35, my emphasis) observation that:

> where arguments do seem to have an effect, this is more often due to their *physical repetition* [in e.g. thought experiments] than to their *semantic content*.

In short, the mental-modeling view fills many gaps in our understand-

ing of STEs that the argument view left unexplained. But, like the argument view, the mental-modeling view also suffers from serious issues. I mention three.[102]

Firstly, it is not obvious that the mental-modeling view explains *all* STEs equally well. The mental-modeling view seems best suited to explain the type of STEs that concern classical mechanical phenomena with which *we*, i.e. *our bodies*, have plenty experience. Clement's Sisyphus is a good example here. But it is hard to see what the mental-modeling view adds to our understanding of more abstract TEs such as Norton's dome or, looking beyond the aforementioned examples, STEs in e.g. modern physics, like the famous EPR-Bell scenario or STEs involving black holes or multiverses, where 'embodied' mental-modeling seems much less directly relevant than reasoning (argumentatively) about abstract ideas. So, at most, the mental-modeling view is adequate only for a sub-set of all STEs.[103]

Secondly, mental models are, in an important sense, *too* subjective to fully account for STEs.[104] Yes, STEs are performed in our minds and so the *performance* of an STE (4.4) is a subjective process. But I have already argued that STEs are much more inter-subjectively stable than this: STEs have *topics*. The mental-modeling view seems limited to describing the subjective experience of STEs but cannot do much to enhance our understanding of the inter-subjectively *stable* aspects — i.e. *the topic* — of STEs (besides explaining the role of the descriptions of STEs). For example, the mental-modeling view is not very suited to help us evaluate what happens when two people *disagree* on the outcome of the same thought experiment because we cannot easily compare the *content* of two different acts of mental-modeling by two different imaginers, hence we cannot easily evaluate their respective 'soundness'. And, ironically, if we

[102] I here broadly follow Meynell (2014)'s discussion of the mental-modeling view.
[103] Meynell (2014) notes that this same objection can be used to argue that the mental-modeling view is not well suited for many thought experiments in philosophy.
[104] Similar objections have been put forward, e.g. by Godfrey-Smith (2007), against Nersessian's "psychologistic" mental-modeling account of *scientific models*.

were to try and evaluate the 'soundness' of some act of mental-modeling, we would presumably do it by reconstructing and comparing *arguments*. In emphasizing the imaginative, experiential and embodied character of the performance of STEs, we should not throw out the baby with the bathwater and *ignore* the epistemic importance of arguments that often underlie STEs.

But perhaps the biggest problem facing the mental-modeling view is that, despite the widespread popularity of the concept of mental models, *it remains unclear what exactly mental models are*. As Meynell (2014, p.4156-7) puts it:

> [T]he exact character of this imagined mental content is up for dispute. [...] [This is] hardly surprising given that among philosophers there is no agreement about the nature of mental content. Nersessian builds her account on recent work in cognitive psychology, arguing that thought experimenting is just one of a group of model-based reasoning practices (2002). But by her own admission, though the cognitive psychology literature is suggestive it is also controversial and diverse (2007). Any position is, to some extent, speculative.

For the case of STEs, the biggest source of dispute lies in two questions: (i) to what extent are mental models *imagistic*, i.e. are mental models composed of *mental imagery*?; and (ii) to what extent are mental models *propositional*, i.e. do they have semantic content?

Different answers to these questions lead to different explanations of the epistemic value of STEs. For example, Miščević (1992, 2007) and Gendler (2004) explained how we gain insight through mental-modeling by appealing to the epistemic value of performing *quasi-observations*; they require the imagined mental content to be *imagistic*, as opposed to propositional, at least to the extent that "the presence of a mental image may play a crucial cognitive role in the formation of the belief in question" (Gendler, 2004, p. 1152), as quoted and discussed in the previous Section. Nersessian (2002, 2007, 2018) however argues explicitly that mental models are *neither* propositional *nor* necessarily imagistic, but rather some

other (perhaps *sui generis*) form of simulative model-based reasoning. But if this is the case, then appealing to *quasi-observations* cannot help us explain the source of the heuristic vlaue and demonstrative force of STEs.

Moreover, if Nersessian is correct and mental models are *not* crucially imagistic, then it becomes hard to see the difference between mental-modeling and 'mere' *abstract reasoning* about spatio-temporal structures and causal relations (i.e. about the *content* of mental models, whatever it is). It becomes hard to see, then, what is distinctively *imaginative* about mental-modeling: it becomes hard to see whether mental-modeling is really an act of *imagination* rather than an act of reason. The essential differences between the argument view and the mental-modeling view appear to dissolve; see Arcangeli (2021) for an elaborate discussion of this issue. And so *still* the mental-modeling view cannot explain epistemic value of imagination in STEs satisfactorily. Meynell (2014, p.4157) soberly concludes:

> Thus while prima facie explanatory, ultimately mental modeling accounts do little to elucidate the content of TEs.

I note that Hacking (1992, p.306) already expressed this same sentiment three decades ago:

> [I]f cognitive science were not so fashionable, I think many of us would feel the same resistance to Nersessian's conceptual models [as to Brown's platonic account]. They are presented as if they were explanatory, but in fact explain nothing.

Oof. We must do better.

4.2.6 Going forward

Before I move on, I briefly recapitulate what I have discussed until now.

I began by elaborating on the two *core questions* concerning STEs:

(I) what *are* STEs?

(II) what, and how, can we *learn* by performing STEs?

I provided lists of many answers to these two questions available in the literature (Section 4.2.1).

I then introduced two STEs at length — Galilei's falling bodies (Section 4.2.2) and Clement's Sisyphus (Section 4.2.2) — and I pointed out important similarities and differences between these two STEs. I then briefly introduced four more (famous) STEs in Section 4.2.2: Newton's bucket, Maxwell's demon, Einstein's photon-box and Norton's dome. With these example-STEs in hand, I then returned to core questions (I) and (II).

Concerning core question (I) (Section 4.2.3), I argued that STEs have three distinct 'ontic components': STEs have (i) *descriptions*, $\mathcal{D}$, which describe (ii) an *imaginary scenario*, $\mathcal{C}$ (the *topic* of the description $\mathcal{D}$), and specify (iii) an epistemic aim, $\mathcal{E}$, that should be achieved by *performing* the STE. To *perform* an STE, I explicated as follows:

> [Performing an STE] Subject S *performs* STE $= \langle \mathcal{D}, \mathcal{C}, \mathcal{E} \rangle$ iff upon engaging with description $\mathcal{D}$, S reasons about imaginary scenario $\mathcal{C}$ (i.e. the topic of $\mathcal{D}$), with epistemic aim $\mathcal{E}$. (4.4)

I then elaborated on each of these three components $\mathcal{D}, \mathcal{C}$ and $\mathcal{E}$. Notably, I described that the imaginary scenarios, $\mathcal{C}$, of STEs give rise to epistemological problems.

Firstly, the imaginary scenarios of STEs contain *idealizations*, they are importantly *incomplete* and they often contain *physically impossible* entities. As such, the imaginary scenarios of STEs are often flat-out physically impossible scenarios. An account of STEs should explain how we can gain scientific knowledge and understanding, even knowledge and understanding *of the natural world*, by engaging with such *impossible* topics.

Secondly, I noted that both the imaginary scenario of an STE and the reasoning-process about this scenario that should lead us to achieve the STEs epistemic aim are typically *under-determined* by the STE's descrip-

tion; there is often significant *variation* in the performance of STEs. And yet we are often able to reach the epistemic aim of STEs in either one of these multiple ways. An account of STEs should explain, then, which features of an STE and which forms of reasoning can vary and which must remain *invariant*. (Getting ahead of myself, I indicated that I shall argue that the only features of an STE's imaginary scenario that must remain invariant are facts about *scientific models*.)

Concerning core question (II) (Section 4.2.4), I notably discussed two important aspects of the epistemic value of STEs: (i) that STEs have more *heuristic value* than non-thought-experimental forms of reasoning, in the sense that we often gain insight *easier* and *more effective* by performing STEs than through other means; and (ii) that STEs have a distinct type of *demonstrative force*, which makes the conclusions drawn from performing STEs particularly convincing or perhaps even uniquely *justified*. I then briefly discussed (and dismissed) the 'paradox of thought experiments', and I discussed the difference between beliefs gained by performing STEs — STE-beliefs — and *quasi-perceptual beliefs*, which were the topic of Chapter 3 of this Thesis.

Finally, I discussed two accounts of STEs that are currently amongst the most popular and long-standing accounts of STEs available in the literature: the *argument view* (Section 4.2.5) and the *mental-modeling view* (Section 4.2.5).

The upshot of the argument view was that this view is 'tantalizingly elegant' with respect to explaining how we can learn about the *world* by performing STEs, namely: just by reasoning through an argument. Moreover, the argument view provides a good methodological rule of thumb for analysing STEs: begin by reconstructing the argument that is conveyed via STE (if there is any). The downside of the argument view was that it is descriptively false and, relatedly, that it cannot explain the heuristic value and demonstrative force of STEs, because it dismisses the psychological role and epistemic value of *imagination* in our performances of STEs.

The upshot of the mental-modeling view was that this view seems much

more descriptively adequate than the argument view: performing STEs often involves setting up and manipulating mental models *à la* Johnson-Laird (1983). The mental-modeling view also explained the psychological and epistemic importance of the *descriptions* of STEs, which it regards as 'user-manuals' for setting up and manipulating mental models, thus guiding our imagination constructively. As such, the mental-modeling view took great steps in explaining the heuristic value and demonstrative force of STEs. The downside of this view, however, was that it seems applicable only to a *subset* of STEs where mental-modeling is paramount, but not to STEs that hinge more strongly on explicit inferential reasoning and argumenting. Moreover, because mental models *are* part of the mental content of individual imaginers, the mental-modeling view cannot account well for the *inter-subjectively stable* aspects of STEs. Additionally, I noted that there is no consensus on *what exactly mental models are*. This led Meynell (2014, p.4157) to the sober conclusion that: "Thus while prima facie explanatory, ultimately mental modeling accounts do little to elucidate the content of TEs."

So much for this recapitulation. I next turn to a topic that lies at the heart of the view of STEs that I propose in this Chapter: the relation between STEs and *fiction*.

4.3 STEs and Fiction

4.3.1 The relation between STEs and fiction

Thought experiments, both scientific and otherwise, have often been related to the concept of *fiction*. The link from literary fiction to thought experiment is perhaps most obvious here. Great works of literary fiction are rarely *just* fiction — they often exemplify, illustrate, amplify and draw our attention to noteworthy features of our actual world and, as such, they enable us to gain insight into the world (Davies, 2007; Swirski, 2007; Camp, 2009; Ichikawa and Jarvis, 2009; Elgin, 2014, 2017). For example, *Anna Karenina* can give us insight into the dynamics of late

19th–century bourgeois imperial Russian society; and "*Flatland* is meant to be a mordant parody of the hypocrisy and closed-mindedness of the Victorian society, and Orwell's *Animal Farm* should be read as an allegory of Stalin's regime" (Sartori, 2023, p.17, fn.25). But the epistemic value of fiction is not limited to literary fiction. It has often been argued that works of *science fiction* can, and should, be understood as thought experiments, precisely because (again) they often have the function to illustrate and amplify some feature of our actual society and, as such, can teach us something about it (Schneider, 2016; Elgin, 2017; Silova, 2020; Wiltsche, 2021; Güzel, 2022).

But it is the other direction of the relation between science and fiction that interests me more: the idea that *scientific thought experiments* can be regarded as *works of fiction*. In philosophy of science, there is a fairly well-entrenched tradition of regarding scientific constructs such as laws of nature, models and theories as *useful fictions*, due notably to the groundbreaking work of Cartwright (1983, 1999); Cartwright and Le Poidevin (1991), which was built upon by (Elgin, 1996, 2010, 2014, 2017), amongst many others, including the references above.[105]

Recently, two accounts of STEs have been provided in the literature that explicitly argue that thought experiments are works of fiction (Meynell, 2014; Sartori, 2023). Importantly, to bolster their proposals, both of these proposals employ the *same* theory of fiction — and I shall employ this same theory of fiction too. I shall soon discuss these two proposals and their underlying theory of fiction at length. But first, I wish to motivate the idea that STEs are works of fiction more thoroughly.

There are at least three inter-related reasons why the concept of fiction is *prima facie* relevant for understanding scientific thought experiments.

[105] Cartwright and Le Poidevin (1991) use the word "fables" rather than "fiction", but the point remains the same. Importantly, their stance is *not* instrumentalist, as they write themselves that philosophers of science "have tended to fall into two camps concerning scientific laws [and other scientific constructs]: either we are realists or we are instrumentalists" (p.55). Neither position, Cartwright and Le Poidevin (1991) argue, can make sense of or do justice to the cognitive efficacy and epistemic value of these scientific constructs — but regarding them as *fables* (i.e. fictions) can.

First and foremost, both works of fiction and the descriptions of STEs *prescribe imaginative engagement* with the topic in question. Both when we read works of fiction, and when we read descriptions of STEs, we are prescribed to perform acts of imagination. This observation, at the very least, suggests that drawing an analogy between STEs and the concept of fiction can be fruitful.

Secondly, both works of fiction and STEs crucially depend on the details of their description to *prompt* and *direct* the acts of imagination that they prescribe. The role of the description is, for both fiction and STEs, multiple: it *prescribes* imaginative engagement with the topic in question, it *describes* the scenario that is to be imagined, and, perhaps most importantly, it *highlights* and draws *attention* to features of this to-be-imagined scenario that are relevant for the epistemic purpose at hand. Perhaps the only feature that distinguishes works of fiction from the description of STEs is that descriptions of STEs *necessarily* specify an epistemic aim that should be achieved by performing the STE, whereas works of fiction do not necessarily have or specify epistemic aims — but they regularly do.

Thirdly, the *fictional worlds* of works of fiction and the imaginary scenarios of STEs are similar in several ways, notably in that they are both inter-subjectively stable, and that both typically contain non-actual, non-existent or even *impossible* entities. Importantly, these aspects are often not epistemically irrelevant but rather epistemically *important*: without imagining these *deliberate falsehoods*, the epistemic aim cannot be gained, at least not as effectively. Orwell's *Animal Farm* is a false description of the real world and yet, precisely because of its false details, it is a particularly poignant illustration of Stalin's regime. Likewise, Maxwell's demon cannot exist in the real world and yet it aptly demonstrates the statistical nature of the Second Law of thermodynamics. Our engagement with works of fiction and STEs can lead us to draw conclusions *effectively* from which the fictional entities are (remarkably) absent.

In short, to regard STEs as works of fiction is to say that (I) descriptions of STEs are works of fiction, and (II) we can learn things about the

world by performing STEs in the same way as we can learn things about the world by engaging with fiction. But this is only the *beginning* of an answer. In order to use a fiction-based account of STEs for *explain* things about STEs, we need a detailed account of fiction. To this I turn next.

4.3.2 Walton's theory of fiction

K.L. Walton's (1990) theory of fiction stands out as one of the most widely embraced and influential contemporary theories of fiction in the literature. Central to Walton's theory is the emphasis on the close relationship between fiction and imagination, which can be distilled into a simple yet profound slogan: *works of fiction prescribe imaginings*. When we engage with a work of fiction, Walton argued, we are obliged to imagine its content. *This obligation* is what makes an object a work of fiction. To provide contrast: if we engage with a work of *non*-fiction, then we are obliged to *believe* its content.

A *work of fiction* is, for Walton, just any concrete object about which there is the convention, in some community, *that* it is a work of fiction and, therefore, that we have to imagine its content.[106] For example, when we read A.C. Doyle's books on Sherlock Holmes, or watch the *Star Wars* films, or look at Salvador Dali's paintings, we understand conventionally that these objects — books, films, paintings — are works of fiction, and thus we understand that we are obliged to imagine, rather than believe, their content.

From the idea that works of fiction prescribe imaginings, Walton's definitions of the two 'core concepts' for any theory of fiction — *fictional truth* and *fictional worlds* — follow straightforwardly. Proposition p is *fictionally true* with respect to a given work of fiction iff anyone (i.e. every member of the relevant community) who were to engage with that work of fiction, would be obliged to imagine that p and would not be obliged

[106] Walton uses the phrase "work of fiction" interchangeably with the term "representation", which is markedly different from *scientific* representation. To avoid confusion, I avoid using the term "representation" in Walton's sense.

to believe that p.[107] The *fictional world* of a work of fiction is just the set[108] of all and only fictional truths of that work of fiction. For the sake of clarity, I provide provisional explications of these core concepts:

> [Work of fiction:] Concrete object w is a *work of fiction* for community C iff there exists a convention in C such that, if any community-member $S \in C$ were to engage with w, then S would be obliged to *imagine* the content of w and S would not be obliged to *believe* the content of w.
>
> [Fictional truth:] Proposition p is *fictionally true* with respect to work of fiction w iff p is part of the content of w.
>
> [Fictional world:] Set of propositions $\mathcal{F} = \{f_1, ..., f_n\}$ is the *fictional world* of work of fiction w iff $\mathcal{F}$ contains all and only fictional truths of w.

(4.5)

I emphasise immediately that there is the common misconception that, if a proposition p is fictionally true, then p is false *simpliciter*. This is wrong. Fictional truth and ordinary truth are logically independent. Sure, fictional truths are *often* false, but they are not necessarily false. To illustrate: it is fictionally true in the fictional world of Sherlock Holmes that the Big Ben stands in London; and it is also true *simpliciter* that the Big Ben stands in London. The fact that one is obliged to imagine that p and one is *not* obliged to believe that p, in virtue of p being expressed in a work of fiction, does not imply that p is false — it only entails that one is not obliged to believe that p in virtue of the work of fiction (but one can of course still come to believe that p in virtue of something *else*); see e.g. (Walton, 1990, §§2.2–2.4); (Frigg and Nguyen, 2022, §3).

[107] Walton (1991) later expressed some reservations about this definition because it seems unable to deal with some subtle issues pertaining to fictional truth, especially for cases of 'fiction-within-fiction'. These reservations are insubstantial for my purpose, so I shall lay them aside (just like every other contemporary philosopher of science that employs Walton's framework). [108] Walton (1990, pp.64-68) resists construing fictional worlds as *sets (or "classes") of propositions* for reasons that are irrelevant for my present purpose. I shall proceed with the simplified assumption that fictional worlds are sets of propositions.

An important question that presents itself from explications (4.5) is: *what, exactly, is the content of a work of fiction?* Some part of the answer that Walton gives to this question is trivial and common to alternative theories of fiction (discussed below), but other parts of Walton's answer are non-trivial and signify core characteristics of Walton's theory, as I shall explain next.[109]

Let me start with the trivial part of Walton's answer. Some part of the content of a work of fiction is provided *directly* by the work of fiction: the text and imagery presented in a work of fiction is, undoubtedly, part of its content. For example: it is written in the books on Sherlock Holmes that Sherlock lives on 221b Baker Street. This is a *fictionally true* proposition about Sherlock Holmes; this proposition is part of the fictional world of Sherlock Holmes. Similarly, the map of Middle Earth that is printed in J.R.R. Tolkien's work of fiction *The Hobbit* is part of the content of *The Hobbit*: the map is a (fictional) representational object that specifies fictional truths about the fictional world of *The Hobbit*.

But now the non-trivial part of the answer. When we engage with a work of fiction, we are generally obliged to imagine *much more* than just those fictional truths that are expressed *directly* in the work of fiction: fictional worlds contain *more* than what is directly expressed in the works of fiction that ground them. In the world of *Sherlock Holmes*, for example, it is fictionally true that Sherlock lives on 221b Baker Street, but surely it is also fictionally true that Sherlock Holmes goes to the toilet sometimes, that he wears underwear, that he has two kidneys and two lungs, that there is blood pumping through his veins at all times, and that he cannot jump 20 meters high or run a marathon in 20 minutes. These fictional

[109] At this point, I should emphasise that we do not only imagine propositions (2.13) when we engage with fiction, contrary to what (4.5) may seem to suggest. We also imagine *actions* (2.31). This is no problem for (4.5), which mentions only propositions, because my explication of *action*-imagination (2.31) demands that, if someone imagines an action, then they dispositionally accept that some proposition (the proposition *that* they perform the action) is possible — *this* proposition is then fictionally true. As such, I hold that we can meet our obligation to imagine that p by imagining an action. Having made this point, I shall focus my discussion predominantly on *propositions* for the sake of convenience.

truths are not expressed *directly* in the work of fiction, but we nonetheless understand that we *are* obliged to imagine these things because it *is* undeniably fictionally true that Sherlock Holmes is a human being and that his physiology is relatively standard (plus he has a knack for abduction and a lack of empathy).

Thus, Walton observed, a distinction arises within the concept of fictional truth, between *direct fictional truths* and *indirect fictional truths* of a work of fiction: *direct fictional truths* are expressed directly in the work of fiction, whereas *indirect fictional truths* are not expressed directly in the work of fiction but nonetheless 'follow from' direct fictional truths, and, hence, are also fictional truths themselves. But *how* indirect fictional truths 'follow from' direct fictional truths is a tricky question; and, in order to provide an answer, Walton had to introduce some new terminology.

To get a grip on *what* exactly the indirect fictional truths of a given work of fiction are, Walton (1990) introduced the concept of *games of make-believe*. In analogy with children's playful prop-oriented games of make-believe such as riding hobby horses, building blanket-castles and playing with toy cars and dolls, a *Waltonian game of make-believe* is a rule-based (and often group-based) act of imagination prompted by and oriented around a prop: the work of fiction. In a Waltonian game of make-believe, a work of fiction prescribes us to imagine specific things (i.e. the direct and indirect fictional truths) according to specific *rules*. The rules that prescribe what to imagine are called the *principles of generation* of the game. These rules are going to give us the indirect fictional truths that we are after.

In terms of concepts explained in the previous Chapter, the principles of generation of a Waltonian game of make-believe are the *constraints* on our imagination: they determine what we should and should not imagine when we engage with a work of fiction, thus they constrain the content of acts of imagination. To make things manageable, I shall interpret these principles of generation as being *propositions*; the principles of generation of a given Waltonian game of make-believe is a set of propositions, de-

noted $\mathcal{R}$, that all members of the relevant community — thus all potential participants of the game of make-believe — adhere to. With this idea in hand, I explicate the concept of a Waltonian *game of make-believe* as follows:

> [Game of make-believe] Subject S's act of imagination A is a *game of make-believe* with concrete object w iff upon engaging with w, due to a convention in S's community C, S performs A such that: A is about the content of w, and A is constrained by set of rules $\mathcal{R} = \{r_1, ..., r_m\}$ that are associated with w, called the game's *principles of generation*. (4.6)

Thus, Waltonian games of make-believe (4.6) are prop-induced, rule-based and *community*-based acts of imagination.

There are always principles of generation at play when we engage with fiction. The most common principles of generation in Waltonian games of make-believe are logical rules of inference, widely shared (scientific) background knowledge and social conventions. Notably, there are always principles of generation involved that oblige us to keep our imagination *as close to the real world as possible*. Even in the most fantastical, otherworldly fiction we generally use at least *some* reality-oriented principles of generation such as the principles of object permanence and the impossibility of two concrete objects occupying the same space. And we typically cannot do without using some quotidian principles of generation too: e.g. the principle that human beings have emotions, desires and ambitions, etc. As such, principles of generation are responsible for, if you will, 'expanding' and 'coloring in' the fictional world of a work of fiction: we employ these principles, deliberately or indeliberately, in our engagement with fiction, because without them these would be very little to imagine *about*; without them, we would not know what a fictional world *is like* beyond its most obvious direct fictional truths. It can be said, then, that the principles of generation associated with a work of fiction determine which

imaginings are *correct* or *incorrect*. They make possible the "coordination of imaginings and concomitantly the identification of which imaginings are authorized" in a specific game of make-believe (Meynell, 2014, p.4158).

Combining (4.6) with the provisional explications of the core concepts in Walton' theory of fiction (4.5), and taking heed of the above-discussed distinction between *direct* and *indirect* fictional truth, the explications of the core concepts in Walton's theory of fiction assume the following form:

Walton's theory of fiction as make-believe:

[Work of fiction:] Concrete object w is a *work of fiction* for community C iff there exist a convention in C such that, if any community-member $S \in C$ were to engage with w, then S would be obliged to play a game of make-believe (4.6) with w, and S would not be obliged to believe the content of w.

[Direct fictional truth:] Proposition p is a *direct fictional truth* about work of fiction w iff p is expressed in w.

[Indirect fictional truth:] Proposition q is an *indirect fictional truth* about work of fiction w for community C iff if any $S \in C$ were to play a game of make-believe (4.6) with w, then S would be obliged to imagine q (because q follows from the direct fictional truths of w and the game's principles of generation $\mathcal{R}$), and q is not a direct fictional truth. (4.7)

[Fictional world:] Set of propositions $\mathcal{F} = \{f_1, ..., f_n\}$ is the *fictional world* of work of fiction w iff $\mathcal{F}$ contains all and only direct and indirect fictional truths of w.

I next make six systematic comments about Walton's theory of fiction that are relevant for my purpose.

First comment. Principles of generation are propositions that constrain the content of our acts of imagination when we play games of make-believe. We may wonder: what is the *propositional attitude* that we adopt

towards these propositions? My choice is *acceptance*. Acceptance seems the adequate propositional attitude for rule-following — the attitude we take towards rules when we intend to act accordingly — and it conveniently coheres with my proposed explication of proposition-imagination (2.13); recall: S imagines that p iff S occurrently accepts that p is τ-possible, for some appropriate modality τ. (Thus: when we *accept* principles of generation for our games of make-believe, we are very close to *imagining* these principles too, which seems appropriate.) The acceptance of the principles of generation can be occurrent or dispositional, just like the propositional attitude of the mental state of imagination can be occurrent (for proposition-imagination 2.14) or dispositional (for action-imagination 2.30), and the principles can be anything from *ad hoc* mandates that are enforced only for a brief moment, to deeply ingrained cultural conventions that are stable inter-generationally.

Second comment. A work of fiction is always a work of fiction *relative to a certain community*. What one community regards a work of fiction, another community may regard a work of non-fiction. The difference is only conventional. The communities mentioned in the explications above may range from a pair of subjects that decide, *impromptu* and for just a moment, to play a game of make-believe with an object they encounter (thus turning that object into a work of fiction for that brief moment), to large-scale communities and even entire cultures that are stable in the long term. All this can be captured by Walton's theory of fiction as make-believe (4.7). Additionally, I mentioned above that Waltonian games of make-believe are often *group-based* acts of imagination: our imaginative engagement with works of fiction can be *shared* and *coordinated* with others — granted that the others are part of the same community, which secures that the others regard the work of fiction indeed as a work of fiction and that they employ the same principles of generation in their acts of imagination, thus securing that everybody imagines the *same* fictional world while playing their game of make-believe.

Third comment. The distinction between direct and indirect fictional

truths is important not only for the sake of formulating a coherent theory of fiction; it is also important because it enables us to explain what it means to *study* and *discover* things *about* a fictional world: we reason about a fictional world with the aim of discovering some of its *indirect* fictional truths, by considering the *direct* fictional truths of a work of fiction and using principles of generation to generate indirect fictional truths. However, while this idea is highly intuitive in the abstract, it is often hard to put into practice, for reasons that I will next expound.

Fourth comment. Principles of generation generate indirect fictional truths — *a lot of them* (Walton, 1990, p.142):

> Fictional truths breed like rabbits. The progeny of even a few primary ones can furnish a small world rather handsomely. We are usually entitled to assume that characters have blood in their veins, just because they are people, even if their blood is never mentioned or described or shown or portrayed. It is fictional in *La Grande Jatte* that the couple strolling in the park eat and sleep and work and play; that they have friends and rivals, ambitions, satisfactions, and disappointments; that they live on a planet that spins on its axis and circles the sun, one with weather and seasons, mountains and oceans, peace and war, industry and agriculture, poverty and plenty; and so on and on and on. All this is implied, in the absence of contrary indications, by the fact that fictionally they are human beings.

Fictional worlds are *massive*, often *infinitely large*, sets of propositions. Notwithstanding, Walton emphasises often, fictional worlds are *incomplete*, in the sense that not every proposition has a truth-value in every fictional world. In the fictional world of Batman, there is no fictional fact of the matter about the exact number of hairs on Batman's body, nor are there fictional truths about whether or not dark matter exists, or about the precise location of the star Betelgeuze: there are simply no principles of generation that prescribe imaginings about these things. To insist on this would, in Walton's words, be "silly" (Walton, 1990, p.174-183).

Here lies the key difference between Walton's theory of fiction and

David Lewis' (1978) famous account of fictional-worlds-as-possible-worlds. Possible worlds are complete: every proposition has a truth-value relative to that possible world. Hence, for Lewis, fictional worlds are complete: in every fictional world of Batman (i.e. every possible world where Batman exists)[110], there is a matter of fictional fact about the number of hairs on Batman's body and about whether dark matter exists or not, even if these fictional truth are 'epistemically inaccessible'; c.f. Kripke (2013). This is not the case for Walton's theory of fiction: some propositions are fictionally true, some are fictionally false[111], and yet many, many other propositions are neither fictionally true nor false — to insist on their truth-value either way is "silly" (Walton, 1990, §4.5). (Getting ahead of myself, I note that this difference is an important reason why contemporary authors on thought experiments and scientific models use *Walton*'s theory of fiction rather than Lewis': the imaginary scenarios of thought experiments and models are *also* incomplete, so Walton's theory is much better suited for describing these scenarios than Lewis' theory is; c.f. (Frigg and Nguyen, 2020, Ch.6).)

Fifth comment. Having said all this, it is important to acknowledge that it is often extremely difficult to identify precisely *which* principles of generation are involved in a given game of make-believe. Principles of generation can be employed not only explicitly but also implicitly. As I discussed in the previous Chapter, the content of our imagination is generally constrained by a hodgepodge of deliberate *and* indeliberate factors (Stuart, 2021). Moreover, Gendler (2007, Ch.7) described how we often switch effortlessly between employing *different* principles within a single act of imagination. Meynell (2014, p.4158-9) argued that we should regard as principles of generation not only deliberately *accepted* proposi-

[110] More precisely: every possible world where the *story* of Batman is "told as known fact" (Lewis, 1978). [111] Whereas fictional *truth* is often discussed explicitly in the literature on fiction, fictional *falsehood* is not. This is a lacuna in theories of fiction, so too in Walton's, further discussion of which is sadly beyond the scope of this Thesis. For my present purpose, it suffices to proceed with a definition of fictional falsehood as the inverse of fictional truth: proposition p is *fictionally false* iff $\neg p$ is fictionally true.

tions that constrain our imagination but also non-propositional constraints on our imagination: psychological capacities, perceptual habits, cognitive tendencies, and implicit "aliefs" — a term coined by Gendler (2008) to denote "associatively linked content that is representational, affective and behavioral". Walton (1990, p.165) concurred:

> If even the flimsiest evidence relation can ground implications [of indirect fictional truths], provided it is reasonably conspicuous, one should expect there to be implications involving no evidence relation at all (neither actual nor believed), but merely a sufficiently salient connection or association of some other sort.

I note that we should also add to Meynell's list of 'non-propositional principles of generation': emotions, moods, hopes, primes, desires and so on, as it is well-known that these types of affective mental content also constrain the content of our imagination.[112] These 'non-propositional principles of generation' are not always easily identifiable, hence not even always clearly *reconstructible* as propositions, and they are uncomfortably *subject-dependent*: fictional truths and fictional worlds do not depend on the whims of individual imaginers, they are more inter-subjectively stable than that. Yet Walton (1990, p.174) himself admits:

> [T]he pouring of the foundations of fictional worlds is no more orderly than the erection of their superstructures; the mechanics of generation are soggy to the core.

Because of these troubles with principles of generation, Walton urges us to distinguish between *two different types of fictional worlds*: (i) fictional worlds *with respect to a work of fiction*, and (ii) fictional worlds *with respect to an individual imaginer*. Suppose, for example, that you are imagining the fictional world of Batman together with a friend: you are playing a shared Waltonian game of make-believe about Batman. It is fictionally true in your game — for *both* of you — that the Joker exists and that he is out to destroy Batman. But what does the Joker look like, what are the

[112] Recall the infamous 'puzzle of imaginative resistance': (Gendler, 2000a; Gendler and Liao, 2016; Tuna, 2020).

fictional truths about its appearance? This may vary wildly between you two, depending for example on which Batman movies each of you has seen in the past: you may imagine the Joker looking like Heath Ledger, Joaquin Phoenix or Jared Leto, while your friend may insist that the Joker looks like Jack Nicholson. Which one of you is *correct*? In some sense, both of you are correct, at least *more* correct than if you were to imagine the Joker looking like Taylor Swift or Beyoncé, or a Smurf. In other words: your personal fictional truth about the appearance of the Joker is *defensible* but not universally shared. But there is also an important sense in which it cannot be the case that both of you are correct: there is only one Joker in the world of Batman, and, given that he is a (fictional) human being, he must have some (single) distinctive appearance.

Therefore, Walton (1990, §1.9) stresses, we must conceptually distinguish between fictional *work worlds* and fictional *game worlds*. On the one hand, there are inter-subjectively stable fictional truths generated by *propositional* principles of generation on which *all* participants of a game of make-believe agree. Examples were the fictional truths about *Batman* that the Joker exists and that he is out to destroy Batman. In Walton's terminology, the collection of all these inter-subjectively stable fictional truths constitute the fictional *work world* of the work of fiction, in this case the work world of *Batman*. Definition:

> **Work world:** The fictional *work world* of work of fiction w is the set of all and only propositions that every participant of a game of make-believe about w is *obliged to imagine*. (4.8)

Thus the *fictional worlds* explicated in (4.7) are fictional *work worlds*.

On the other hand, there are the defensible subjective fictional truths imagined by an individual participant of a given game of make-believe. Examples were fictional truths about the detailed appearance of the Joker — some of these fictional truths are fictional truths *for me* but not for you, and vice versa. The collection of all the fictional truths that some individual participant of a game of make-believe *actually* imagines consti-

tute the *game world* of the game of make-believe that *they* are currently playing. Definition:

> **Game world:** The fictional *game world* of a participant of a game of make-believe about work of fiction w is the set of all and only propositions that that participant *actually imagines*. (4.9)

Work worlds and game worlds *should* partially overlap: to aim for overlap is precisely what it means to *participate* in a game of make-believe about a work of fiction. But generally neither will be a subset of the other. Work worlds contain infinitely many fictional truths (which "breed like rabbits"),[113] while game worlds are always finite (because one always actually imagines only a finite amount of things). Work worlds will generally be much larger than game worlds, yet game worlds will always contain some fictional truths that are *not* part of a work world, e.g. the 'subjective' fictional truths about the Joker's appearance that others do not accept.

The distinction between work worlds and game worlds is epistemologically important because it enables us to evaluate the *correctness* of the imaginings of a participant of a game of make-believe: roughly speaking, you play a *correct*, or 'authorised', game of make-believe about a given work of fiction iff your game world maximally overlaps with the work world of that work of fiction. *This* can be used for epistemological analysis of STEs, as we will see below.

Sixth comment. Walton (1990, p.40) emphasises that fictional *game worlds* are subject-dependent, but fictional *work worlds* are not: in an important sense, fictional work worlds 'exist' independently of whether someone imagines them, they enjoy some degree of 'objective' existence:

[113] Although we cannot *actually* imagine infinitely many things, we can be *obliged* to imagine infinitely many things. If one feels uncomfortable with an obligation to do infinitely many things, then it may help if they add a 'relevance-parameter' in the conditional on the right-hand side of the explication of *work of fiction* (4.5): "if any community-member $S \in C$ were to engage with w, then S would be obliged to imagine the *relevant* content of w [...]", where 'relevance' is relative to the current topic and context of the game of make-believe that S is playing.

> The role of props in generating fictional truths is enormously important. They give fictional worlds and their contents a kind of objectivity, an independence from cognizers and their experiences which contributes much to the excitement of our adventures with them. [...] Fictional worlds, like reality, are "out there," to be investigated and explored if we choose and to the extent that we are able. To dismiss them as "figments of people's imaginations" would be to insult and underestimate them.

Meynell (Meynell, 2021, p.4) noted recently:

> The genius of Walton is that he offers a theory [of fiction] [...] that depends on the imagination without ever trying to enter the heads of imaginers.

So much for Walton's theory of fiction as make-believe. I shall next argue that Walton's theory of fiction is an incredibly suitable conceptual framework to use for understanding STEs. I am not the first who argued for this. Meynell (2014) was first: she explicitly proposed an account of thought experiments (both scientific and philosophical) that is explicitly based on Walton's theory of fiction as make-believe (4.7). With her proposed account of thought experiments, Meynell (2014) argued that she improves on both the *argument view* of TEs and on the *mental-modeling* view of TEs, finding a golden 'middle way' between these two views that incorporates their respective strengths and avoids their weaknesses. Let us see how.

4.3.3 Meynell's Waltonian account of STEs

To introduce the general idea of Meynell's Waltonian account of thought experiments, I give the word to Meynell (2014, p.4161-2) herself:

> When we take Walton's approach and apply it specifically to TEs we get something like the following: [descriptions of] TEs are narratives that are created to prompt their readers to imagine specific fictional worlds, as kinds of situational set-ups that, when you "run," "perform," or simply imagine them, lead to specific results. Often one is

> to imagine oneself doing or experiencing something in this fictional world. It is this relationship between set up and results and the concomitant analogy with real experiment that gives TEs their experimental character, though the many narratives that we call TEs can be more or less experimental. One is directed to attend to the result — what ultimately happens in the imagined scenario [...] — or one's own thoughts or feelings as a participant/observer in this scenario [...] or even the principles of generation that one draws from in imagining the scenario [...]. The content of a TE is determined by the words (and any associated image) [of that TE's description] together with principles of generation.

And, specifically on *scientific* thought experiments, Meynell continues (*ibid.*):

> These principles [of generation] can be stipulated in the TE and include widely understood conventions, background beliefs (both tacit and explicit) and aliefs, habits of mind, basic cognitive and perceptual capacities and expectations. Some of them may be specific to a particular community or discipline. So, for instance, the TEs of physics may make use of a different set of principles than those of philosophy. [...] Fictional worlds are incomplete, with fuzzy boundaries but nonetheless for a well-designed TE there are a distinct set of fictional truths that a reader imagines when reading a TE properly. These fictional worlds include fictional truths that are also actually true—a feature that allows TEs to provide insights into the real world. [...] In the sciences, the insights typically concern what is or is not the case in the real world; for instance, imagining dropping cannon balls and musket balls off buildings gives us insight into free fall. I use "insight" here as a deliberately vague term. This keeps it open to the many different things that various TEs do, from showing inconsistencies within a theory (Brown, 1991, pp.34-36), to making a theory appear plausible (Brown, 1991, pp.36-38), to clarifying a conceptual schema (Kuhn, 1977), to justifying an existing human made-system (Reiss, 2012), to revealing background assumptions (Gendler, 1998) or questioning them (Camilleri, 2014a). Many scientific TEs provide grounds for accepting or rejecting specific claims.

For the sake of clarity, I formulate Meynell's Waltonian account of STEs in terms of my definition of an STE (4.3). In terms of Meynell's Waltonian account of STEs, a *scientific thought experiment* STE is an ordered triple:

$$\mathsf{STE} = \langle \mathcal{D}, \mathcal{F}, \mathcal{E} \rangle, \tag{4.10}$$

consisting of a *description* $\mathcal{D}$, a *Waltonian fictional world* $\mathcal{F}$ (4.7), and an *epistemic aim* $\mathcal{E}$. Thus, what it means to *perform* an STE, is the following:

> Subject S *performs* $\mathsf{STE} = \langle \mathcal{D}, \mathcal{F}, \mathcal{E} \rangle$ iff upon engaging with description $\mathcal{D}$, S plays a game of make-believe (4.6) with $\mathcal{D}$ and reasons about its fictional world $\mathcal{F}$, with epistemic aim $\mathcal{E}$. (4.11)

Such a Waltonian account of STEs, Meynell argues (and I agree), greatly increases our understanding of thought experiments, and it improves on the argument view and mental-modeling view of TEs in various respects. In particular, Meynell (2014, p.4163-7), see also Meynell (2018), argues that her account improves on the argument view because her account explains (i) why TEs are often accompanied by pictures and diagrams: because they are *props* for imagining fictional worlds, and pictures and diagrams are often more efficient than words in specifying direct and indirect f-truths (in the spirit of: 'a picture is worth a thousand words'); (ii) the ubiquity of experiential and perceptual language in TEs and our discussions of them: because the games of make-believe that we play with works of fiction, hence with thought-experimental scenarios, demand *active*, i.e. *experiential*, imaginative participation; and (iii) the importance of the imaginative character of TEs: the imagined content of thought experiments "and the ways in which they are produced and provide insight simply does not always have a propositional or argumentative form" (Meynell, 2014, p.4165), *contra* the argument view and *gratia* the mental-modeling view.

Meynell also argues that her Waltonian account of STEs improves on the mental-modeling view because her account incorporates the insights of

the mental-modeling view — active imaginative participation in games of make-believe often involves mental manipulation of imaginary scenarios and, hence, straightforwardly *involves* mental-modeling (and argumenting, for that matter) — but it can explain the important role of imagination in STEs in a way that does not depend *only* on mental models. Mental models are part of the content of individuals' mental states (of imagination). As such, they are, in Waltonian terms, part of the STE's *fictional game world*. Alongside this, the Waltonian account of STEs enables us to get a grip also on the inter-subjectively stable imaginary scenario of an STE: the *fictional work world* $\mathcal{F}$ of the STE.

Along this same line of thought, using the Waltonian distinction between work worlds and game worlds, Meynell's account of STEs also explains what it means for multiple people to perform the *same* STE and what occurs when people *disagree* on the outcome of a given STE, which is something that neither the mental-modeling view nor the argument view can explain well (Meynell, 2014, p.4166):

> We perform the same TE when we imagine two game worlds which share the same fictional truths as the work world of a TE. The fictional truths of the work world are what confer identity, which allows that rather different descriptions of a TE (for instance, the many different translations of Lucretius' TE)[114] are still importantly the same TE. Differences of interpretation arise when two different people construct two different game worlds on the basis of the same work world in such a way that this affects the resulting insight (as for instance, Mach and Newton's alternative accounts of Newton's rotating bucket TE). If the difference rests on the fact that one thought experimenter applies the wrong principles of generation then the difference is trivial, it is simply a case of misrepresentation. However, some times two [...] game worlds might be constructed with importantly differ-

[114] "In *De Rerum Natura*, Lucretius presented a thought experiment to show that space is infinite. We imagine ourselves near the alleged edge of space; we throw a spear; we see it either sail through the 'edge' or we see it bounce back. In the former case the 'edge' isn't the edge, after all. In the latter case, there must be something beyond the 'edge' that repelled the spear. Either way, the 'edge' isn't really an edge of space, after all. So space is infinite." (Brown, 2007, p.155).

> ent fictional content and results. Bracketing off those TEs for which the insight involved actually rests on there being a diversity of results, in such cases we have reason to suspect that the TE may be poorly constructed. Perhaps it is unclear which are the appropriate principles of generation, or perhaps there is some background belief, alief, or convention that is assumed by the TE's author and a number of members of the community that is in fact controversial.[...] [This] directs us to pick apart the principles of generation and identify the source of the disagreement and then assess the nature of this source, demanding whether the thought experimenters in question accept the relevant principles or if they reflect default psychological tendencies or implicit biases that the thought experimenter consciously disavows.

In short, Meynell's Waltonian account of STEs explains a great deal about the imaginary scenarios of STEs and our engagement with them, by reconstructing the inter-subjectively stable imaginary scenarios of STEs as *fictional work worlds* and our imaginative performance of STEs as *fictional game worlds*. This provides, in principle, a straightforward analytic strategy for reconstructing and evaluating the *performance* of STEs. First, consider the description of the STE (regarded as a Waltonian work of fiction) and the imaginings it prescribes, i.e. reconstruct the fictional *work world* of the STEs. Then, identify the principles of generation that individual thought experimenters employs while performing STEs, i.e. reconstruct the *game worlds* of individual thought experimenters. Finally, by comparing these game worlds with the work world of the STE, we can evaluate whether a given thought experimenter performed an STE correctly or incorrectly, and we can understand what happens when two thought-experimenters disagree on the outcome of the same STE.

However, by Meynell's own admission, there is a clear limitation to her account: "for while a Waltonian approach elucidates the content of TEs it does not provide substantive rules for assessing that content" and the insight (or justification therefore) it purportedly provides about the natural world (Meynell, 2014, p.4163). A Waltonian account of STEs enables us

to reconstruct the imaginary scenarios of STEs and our engagement with them, but it does not straightforwardly enable us to *epistemologically evaluate* STEs whose epistemic aims concern the natural world. We have a principled method for reconstructing imaginary scenarios of STEs as Waltonian fictional worlds, but we do not *thereby* know how these fictional worlds relate to the natural world. Meynell writes (*ibid.*, p.4163-4):

> All we can really say at this point is that a complete method for epistemically evaluating any given TE will be a two-step process, shaped by the distinction discussed above between the content and epistemological function of TEs. [...] First, because TEs are narratives with the function of prompting and guiding imaginings of a fictional world we must analyse this content and identify the principles that generate it. Second, because these imaginings have a point — they are intended to produce some kind of insight — we need to identify the character of this insight and the evidence or justification that the fictional world provides for it. The Waltonian analysis is sufficient to achieve the first step, but offers only some of the basic tools required to achieve the second.

Meynell submits that a complete method for epistemically evaluating any given STE will be a two-step process: the first step being the analysis of the fictional world of an STE, and the second step being the identification and evaluation of the insight that the STE is supposed to provide.[115] The Waltonian conceptual framework is sufficient to handle the first step, but it cannot handle the second.

Before I continue, it is important to note that Meynell's two-step epistemological framework for STEs is *not* the same process as *the two-step process for quasi-perceptual beliefs* (3.9) that I introduced in the previous

[115] Concerning this two-step method for epistemologically analyzing STEs, Meynell notes that it "is not a novel suggestion. Häggqvist (2009, esp. p.62) and Miščević (2007, esp. p.199) both endorse types of two step analyses of TEs with steps similar to my own, an imagining step and a generalization or argumentation step." I note (somewhat begrudgingly) that Sartori (2023) falsely claims that Meynell does not consider this topic at all: it is, in fact, one of the main take-aways from Meynell's discussion.

Chapter of this Thesis. I have already argued why this is the case in Section 4.2.4, but it will be instructive to review the difference between the two-step process for quasi-perceptual beliefs and particularly Meynell's mentioned two-step epistemological framework for STEs.

The two-step process for quasi-perceptual beliefs (3.9) was a reconstruction of the *formation* of quasi-perceptual beliefs: the first step being a quasi-perceptual process, upon which a proposition comes to mind; the second step being an inferential process concerning the accuracy of our imaginings, upon which we adopt the attitude of belief to the proposition that came to mind in the first step. Meynell's two-step epistemological framework for STEs is different: it does not aim to reconstruct *how* we form beliefs on the basis of STEs but, rather, it is an epistemological method for *evaluating* these beliefs. The first step of this method concerns the reconstruction of the content of the fictional scenario of the STE, and of the reasoning process of some performer of the STEs *about* this fictional scenario, aiming to identify the insight that they gained *about the fictional world* and the principles of generation that led them there. The second step concerns identifying the insight that the STE-performer purportedly gained by performing the STE, which may (or may not) be insight *about the natural world*, and to evaluate the evidence or justification that the fictional world of the STE provides for it.

The first step of this two-step process can be achieved in the Waltonian conceptual framework that Meynell employs for her account of STEs. The second step can only be achieved once we have a detailed account of *scientific representation*. If STEs are supposed to teach us anything about the natural world, then the fictional worlds of STEs must *represent* the natural world. Accounting for this representation-relation between the fictional worlds of STEs and the natural world is not a trivial affair. Scientific representation has long been analysed in terms of *similarity*, *resemblance* or *isomorphism*: roughly speaking, A accurately represents B only if A is sufficiently *similar* to B. This requires that the representans (that which does the representing) and the representandum (that which is

represented) *share properties*. But this idea, that representans and representandum must be sufficiently similar, hence must share many properties, does not work well for the case of STEs, for two main reasons.

Firstly, as I discussed several times, the fictional worlds of STEs contain many idealizations, abstractions and non-existent or even *impossible* entities — the fictional worlds of STEs are often *physically impossible* scenarios — and hence need not be *similar* to the natural world at all, in the sense that they share *many* properties. And yet we can learn about the natural world by engaging with the *impossible* fictional worlds of STEs. So, to evaluate STEs, we require an account of representation that does not hinge on similarity or a related concept; c.f. Frigg and Nguyen (2021a).

Secondly, representation is often considered to be a relation between two *concrete* objects. But fictional worlds of STEs are not concrete objects. Hence it is a genuine (metaphysical) question in which sense the fictional world of STEs can be said to have properties *at all*, let alone how they *represent*; c.f. Salis (2021); Salis et al. (2020).

Sartori (2023) argued that a recently-developed account of scientific representation can deal with both of these problems and, therefore, is perfectly suited for analyzing STEs: the DEKI-account of representation introduced by Frigg and Nguyen (2016, 2017a,b, 2018, 2020).[116] To this suggestion I turn next.

4.3.4 The Walton-DEKI account of STEs

Following Meynell (2014), Sartori (2023) proposed an account of STEs that is explicitly built on Walton's theory of fiction. In order to improve on Meynell's account, Sartori additionally employs for his account of STEs an account of scientific representation, the DEKI-account of representation, to account for the relation between the fictional worlds of STEs and the natural world. I next introduce this DEKI-account in my own words.

[116] I made this same suggestion in early presentations of my research, e.g. Rijken (2020, 2021a,b), and in a manuscript about my proposed *fiction view of STEs* submitted to the *BJPS* in 2019, which formed the basis of the current Chapter.

The DEKI-account of scientific representation

'DEKI' is an acronym for *Denotation*, *Exemplification*, *Keying-Up* and *Imputation* (Frigg and Nguyen, 2016, 2017a,b, 2018, 2020). In a nutshell, according to the DEKI-account of scientific representation, representans A *accurately represents* representandum B iff the following four conditions hold:

(i) A *denotes* B,

(ii) A *exemplifies* set of properties $\mathcal{P}$ of A,

(iii) there exists a *translation key*, $\mathcal{P} \rightarrow \mathcal{Q}$, that maps $\mathcal{P}$ to a different set of properties $\mathcal{Q}$,

(iv) set of properties $\mathcal{Q}$ can be *imputed* on B.

Condition (i), *denotation*,[117] is a necessary condition for being able to talk of *representation* at all (additionally, it secures that representation is an *asymmetric* relation; recall Chapter 2; Section 2.2). Conditions (ii), (iii) and (iv) can be used to distinguish *accurate* from *inaccurate* representation, as follows. If (i)–(iv) are true for some relation between two objects A and B, then A *accurately represents* B; if either (ii) or (iii) or (iv) is only partially true, then A *inaccurately represents* B.

The DEKI account of representation is designed specifically to account for cases of representation where representans and representandum do *not* share many properties. Frigg and Nguyen's (who proposed the DEKI-account) favorite example is the MONIAC (Monetary National Income Analogue Computer), which is a 2-meter high machine consisting of plastic tanks and pipes containing water; see e.g. (Frigg and Nguyen, 2020, §8.1). The MONIAC was built to *represent* England's national economy; see e.g. Bissell (2007) for a historical background. See Figure 4.9 for a picture of a MONIAC in London.

The MONIAC is an unusual construction that, on first sight, has nothing to do with England's economy. The two cannot be said to *resemble*

[117] Denotation is the same as *reference*.

Figure 4.9: A MONIAC in London. (Image courtesy of wikipedia.com.)

each other; they share few properties, if any. The MONIAC functions due to water-pumps and gravity, neither of which are a driving force of England's economy. And yet, as is evident from its popularity, the MONIAC managed to represent England's economy quite aptly. The amount of water in a certain tank represents the amount of money available to a certain system, such as the healthcare system or education system, which are represented as water-tanks. Due to gravity, water *falls* down in the machine, which represents the top-down *spending* of money by governmental institutes. A water-pump pumping the water back up represents taxation. And so on.

In terms of the DEKI-account of representation, using its conditions (i)–(iv), an analysis of the representation-relation between the MONIAC and England's economy would proceed along the following lines.

(i) Denotation. The MONIAC was *built* to represent England's economy. This is explicitly *written* on the MONIAC's description plaque, for

example. Hence it is true that the MONIAC *denotes* England's economy.

(ii) Exemplification. The MONIAC explicitly *exemplifies* a set of properties, such as the amount of water in a given tank, the size of water-tanks and the direction of the flow of water. This is the set of properties $\mathcal{P}$ exemplified by the representandum.[118]

(iii) Keying-up. As I described above, there exists a translation key that maps the set of properties $\mathcal{P}$ onto a different set of properties $\mathcal{Q}$: e.g. sizes of water-tanks map onto amounts of available money for a given governmental system, direction of flow of water maps onto the direction of spending of money, and so on. All this has been explicitly written, e.g. in the description-plaques accompanying the MONIACs.

(iv) Exemplification. Finally, it should be the case that England's economy truly *has* properties $\mathcal{Q}$, such that the translation key 'works' for this particular representation-relation and we can impute properties $\mathcal{Q}$ onto it. Evidently this is the case for the relation between the MONIAC and the parts of England's economy that the MONIAC denotes: the MONIAC accurately represents (a part of) England's economy.

There is much more to be said about this account of representation, most of which is unfortunately beyond the scope of this Thesis. What matters for the present purpose is how we can use this account to analyse the representation-relation between the fictional worlds of scientific thought experiments and the natural world. To this I turn next.

Applying DEKI to STEs

Two questions present themselves with particular urgency when we want to apply the DEKI-account directly to STEs. First: *what* does the denoting (Thomasson, 2020; Friend, 2020)? Second: in which sense do fictional worlds of STEs *exemplify* properties (Frigg and Nguyen, 2020; Salis et al., 2020)?

[118] Note that the MONIAC is also replete with *relations*, as e.g. money flows from one tank to another. Relations can also be represented. For the sake of expediency, I limit my attention to *properties* in my discussion of DEKI-representation (just like Sartori does).

I begin with the first question: *what* does the denoting? As Frigg and Nguyen (2020, p.186) themselves indicate: the key to using the DEKI-account for understanding how fictional worlds — so too the fictional worlds of STEs — can represent the natural world is "to look at the role of descriptions". Descriptions are concrete objects which are perfectly capable of denoting. The example of the MONIAC was a case in point. This idea can be carried over to STEs straightforwardly.

On to the second question: in which sense do fictional worlds of STEs *exemplify* properties? The answer to this question is slightly more complex than the answer to the first problem. Fictional worlds are sets of propositions (4.7), and therefore, strictly speaking, fictional worlds only have the type of properties that sets of propositions do. But this is a problem because *sets of propositions do not exemplify the type of properties that concrete objects like the MONIAC do.* It cannot be said that the imaginary water that 'inhabits' the fictional world of Newton's bucket exemplifies the property of having a concave surface in the same way as it can be said that *real* water in the MONIAC exemplifies the property of flowing in a certain direction. It may seem that we have here encountered a deep metaphysical issue pertaining to fiction; it has often been interpreted as such. But I submit there is a rather natural solution available, which is endorsed by Frigg and Nguyen (2020) themselves: we must appeal to *imagined exemplification.*[119]

In the context of Walton's theory of fiction (4.7), the idea of imagined exemplification — call it *i-exemplification* — is straightforwardly explicated: a fictional work world *i-exemplifies* some property iff *if we were to play a game of make-believe about that fictional world, then we would be obliged to imagine some entity having that property.* Equivalently:[120]

[I-exemplification] The fictional work world $\mathcal{F}$ of work of fiction w *i-exemplifies* property P iff it is fictionally true about w that there exists some entity $\varepsilon(P)$ that exemplifies property P. (4.12)

[119] See also Kripke (2013). [120] *Mutatis mutandis* for i-exemplification of *relations*; recall footnote 118, p.270.

To illustrate: the fictional character Sherlock Holmes i-exemplifies a knack for abduction; likewise, the water in Newton's bucket i-exemplifies a concave surface.

Incorporating Meynell's Waltonian account of STEs (4.10) with the DEKI-account of representation explained above, and substituting therein the notion of exemplification with *i-exemplification*, we obtain the following conditions for when a scientific thought experiment *accurately represents* its *target system* T (if there is any), about which performing the STE should help us achieve some epistemic aim $\mathcal{E}(T)$:

> Scientific thought experiment $\mathsf{STE} = \langle \mathcal{D}, \mathcal{F}, \mathcal{E}(T) \rangle$ *accurately represents* target system T iff:
> (i) description $\mathcal{D}$ *denotes* T by specifying epistemic aim $\mathcal{E}(T)$,
> (ii) fictional world $\mathcal{F}$ *i-exemplifies* set of properties $\mathcal{P}$,
> (iii) there exists a *translation key*, $\mathcal{P} \rightarrow \mathcal{Q}$, that maps $\mathcal{P}$ to a different set of properties $\mathcal{Q}$, and
> (iv) set of properties $\mathcal{Q}$ can be *imputed* on T such that $\mathcal{E}(T)$ can be achieved. (4.13)

I next review the account of STEs put forward by Sartori (2023).

Sartori's Walton-DEKI account of STEs

In a nutshell, Sartori (2023) proposes an account of STEs that is identical to Meynell's account of STEs explicated in (4.10) and (4.11), and adds to it the DEKI-account of representation applied to STEs (4.13).

While I mark Sartori's proposal as a step in the right direction, I note that Sartori uses a much less detailed account of Walton's theory of fiction as make-believe (4.7) than Meynell does. Sartori does not, for example, distinguish between game worlds and work worlds, which is a problem because this distinction does a lot of work in Walton's theory of fiction and in Meynell's account of STEs. Moreover, Sartori draws a strong analogy between thought experiments and *real* experiments, which analogy is

intuitive in principle (and which has been drawn many times, see notably Sorensen (1992); Arcangeli (2018)), but the particular analogy that Sartori draws is rather moot with respect to increasing our understanding of STEs, as Sartori himself partly acknowledges and as I shall argue below.

In analogy with Campbell's (1957)'s bifurcated account of 'real', material scientific experiments, which distinguishes between internally valid and externally valid results of a scientific experiment, Sartori (2023) distinguishes between *internally* and *externally valid* performances of STEs.[121] Explications (reformulated using the concepts of this Thesis):

> [Internal validity] Subject S's performance of an STE (4.11) is *internally valid* iff upon performing STE, S obtains the true belief that the fictional world of STE i-exemplifies property P.
>
> [External validity] Subject S's performance of an STE (4.11) is *externally valid* with respect to real-world target system T iff on the basis of an internally valid performance of STE, and on the basis of the representation-relation between STE and its target system T, S obtains a true belief about T. (4.14)

This distinction (4.14) makes intuitive sense, but it does surprisingly little explanatory work. Indeed, Sartori's distinction between internal and external validity of STEs seems a mere *relabeling* of Meynell's two-step method for epistemologically evaluating STEs discussed in the previous Section.

What Sartori added to Meynell's two-step method, is an explicit account of representation that can take step two (the representational step). But the way Sartori cashes this out — just by introducing the DEKI-account of representation — achieves very little. What is still missing, as I shall argue next, is a principled *method* for reconstructing the fictional worlds of STEs. Without this, it remains unclear to *what* we ought to

[121] Sartori himself talks of *internally and externally valid STEs results*, but he remains vague on what STEs or their results are composed of, and so I interpret his distinction as applying to the *performance* of STEs.

apply the DEKI-account of representation. Sartori introduced the DEKI-account of representation, but he gives no instructions for how to apply it. (Again, getting ahead of myself: reconstructing the imaginary scenario of STEs as *scientific models* does precisely this.) To illustrate this lacuna in Sartori's account, I next review one example STE that Sartori discusses: Maxwell's demon.

Sartori argues that Maxwell's demon is an STE that hinges on *internal* validity: if one performs Maxwell's demon and obtains the belief that the Second Law of thermodynamics is violated *in the imaginary scenario*, then the STE is internally valid. The external valid result would be to conclude that the Second Law of thermodynamics is in reality a *statistical* law, which conclusion, Sartori (2023, p.10) notes, " does not immediately follow — there are no such things as Maxwellian demons in the world." Sartori does not note any connection between the internal and externally valid results of Maxwell's demon, nor does he point towards a way of making this connection using his proposed account.

To me, this analysis of Maxwell's demon is rather underwhelming. We already *knew* that the Second Law of thermodynamics is violated in the imaginary scenario of the STE — *this* 'internal' result is essentially a background assumption of Maxwell's STE. The real point of the STE is to show this result in an imaginary scenario that is designed in such a way that should *help* us achieve some externally valid result, i.e. in a way that *suggests* that the result is externally valid: to conclude that the Second Law of thermodynamics *is* statistical in nature. An account of STEs should point out the aspects of the imaginary scenario that help us in doing so, but Sartori's account remains silent on this matter. We do not know *what* to apply DEKI to. Sartori (2023, p.23) acknowledges that:

> There is no ready-made recipe to determine whether a TE is externally valid. In this TEs are like other surrogative system such as models or [material experiments] aiming at extrapolation.

Moreover, Sartori (2023, p.24) himself questions the usefulness of the distinction (4.14) between internally and externally valid performances of

STEs:

> It is important to clarify that [distinction between internal and external validity] does not imply that the imaginary scenario and its representational function have nothing to do with each other. When it comes to actually constructing TEs, scientists will obviously make considerations of empirical and theoretical nature. Therefore, they will aim from the start at accurately representing something in the world. This does not undermine [distinction between internal and external validity] because the two types of validity, although being intertwined in practice, remain conceptually distinct.

Sartori questions the practical usefulness of the distinction (4.14), but he nonetheless insists that the two remain conceptually distinct. I beg to differ: internally and externally valid performances are not conceptually distinct, they are conceptually *inter-dependent*. The explications of internal and external validity (4.14) that I provided show clearly that external validity *implies* internal validity.

Besides this, it seems that, in practice, *the performance of an STE is internally valid only if the thought-experimenter employs the principles of generation that are necessary for establishing externally valid results.* The principles of generation that determine external validity are generally the *same* principles that determine internal validity. We imagine the water in Newton's bucket rising along the edge *because* we believe that it would occur like that in the real world; we imagine Galilei's tied-together falling bodies fall at the same rate that the bodies would fall individually and unconjoined *because* we believe that this is what occurs in the real world.

This point bears repeating. Even though the imaginary scenarios of STEs often contain many non-existent and impossible entities, the principles of generation that govern these imaginary worlds — and which determine both the internal validity *and* the external validity of STE performances — *are* reality-oriented principles of generation. Thus the same principles that determine external validity of STEs also determine their internal validity. (This again, strengthens the analogy between STEs and

scientific models.) I conclude that the distinction between internal and external validity does not increase our understanding of STEs in the ways that Sartori had envisioned.

Sartori promised to improve on Meynell's account, but he succeeded only partly. Sartori discusses at length how his account 'stabilises the debate' between notably the argument view and the mental-modeling view of STEs by incorporating the important insights from both accounts. While this is true, I content that Meynell had already achieved this by introducing her account, as I discussed in Section 4.3.3. Sartori's introduction of the DEKI-account of representation-by-fictional-worlds has demonstrated that it is *possible* to epistemologically evaluate STEs within a Waltonian framework. At the very least, it is a proof of concept that there exist accounts of scientific representation that are suitable for this task. But the way in which Sartori brings this into practice — by distinguishing between internal and external validity of STE performances — is not very fruitful as it does not provide us with a general *method* for applying DEKI to STEs, hence he does not provide a method for epistemologically evaluating STEs that we did not already have.

I suggest that there is a natural way of improving on both Meynell's and Sartori's account of STEs: we should reconstruct the fictional worlds of STEs as *scientific models*. This, as I shall argue in the next Sections, makes sense of the representational function of the fictional world of STEs: the fictional worlds of STEs represent the world *just like* scientific models represent the world, for the simple reason that the fictional worlds of STEs *are* scientific models. More precisely: *the features of the fictional worlds of STEs that are relevant for its representation-relation (and which must remain invariant in our varying performances of STEs) are facts about scientific models*. Scientific models, in other words, *underlie* the fictional worlds of STEs. *This* suggestion provides an explicit method for analysing and evaluating STEs, both with respect to the reconstruction of their imaginary scenarios (and the 'internally valid results' about them) and with respect to their representational function (and the 'externally

valid results' obtained on the basis of this function): reconstruct the scientific models that underlie the fictional worlds of STEs and evaluate the representation-relation between *those* and the world.

I next motivate the connection between STEs and scientific models more thoroughly.

4.4 STEs and Models

4.4.1 The relation between STEs and models

An important yet underappreciated way of understanding STEs focuses on the close relation between STEs and *scientific models*. Reinier and Burko (2003, p.367) noted carefully that "occassionally the dividing line between a TE and a theoretical model may be blurred." Going further, Morgan (2002, 2004) described the practice of thought-experimenting *as* scientific model-building, and vice versa; and Markie (2005) argued that many theoretical models (in economics) *are* thought experiments. Along similar lines, Cooper (2005) argued that thought experiments pose what-if questions that we answer by constructing models — often, though not always, by constructing *mental* models. From a slightly different angle, Boniolo (1997) proposed a unified theory of thought experiments and scientific models by considering both as fictional "as-if" constructions *á la* Vaihinger (1924). More recently, Arcangeli (2010, 2017) and Salis and Frigg (2020) similarly argued that thought-experimenting and model-based reasoning involve the use of (propositional) imagination in highly similar or even identical ways, and El Skaf and Stuart (2023) discussed the relation between thought experiments and scientific models in broad outlines, indicating many promising directions for future research.

I note that the connection between STEs and scientific models had been forged long before all this, in a revealing and curiously overlooked passage from P. Suppes' (1960) seminal article on scientific models (Suppes, 1960, pp.296-7), which appeared more than half a century ago:

> An important use of models in the empirical sciences is in the construction of Gedanken experiments. A Gedanken experiment is given precision and clarity by characterizing a model of the theory which realizes it. [...] It is my own opinion that a more exact use of the theory of models in the discussion of Gedanken experiments would often be of value in various branches of empirical science."

Suppes called for a more exact use of the theory of models in the analysis of STEs. I shall respond to this call, albeit not in the way that Suppes called for.

Drawing a link between STEs and scientific models is promising in many ways. Most promising is the link between STEs and *theoretical* models, i.e. models that exist not as concrete systems but rather exist only 'in virtue of their description'; think e.g. of Bohr's model of the Hydrogen atom or the model of the ideal pendulum (discussed in the next Section). This is the connection that I shall argue for. Just like theoretical models, STEs too exist only 'in virtue of their description'. Moreover, both theoretical models and the fictional worlds of STEs contain idealizations and often non-existent or even impossible entities — and these features are not epistemically irrelevant but rather epistemically *crucial* in both cases. Additionally, theoretical models and STEs enable us to learn about the world by studying, respectively, a model system or the fictional world of an STE. In other words, both theoretical models and STEs enable *surrogative reasoning*: we can learn indirectly about some real-world target system *by* reasoning about models or performing STEs. The concept of representation is of course crucial here: surrogative reasoning works only if the surrogate system — the model system or the fictional world of an STE — accurately represents the target system.

One difference between theoretical models and STEs that springs to mind is that there is generally much more mathematics involved in descriptions of theoretical models than there is in the descriptions of STEs. This is mostly a problem for STEs in physics. Neither models nor STEs in e.g. biology are generally math-heavy at all. Even (formal) models in economics

are not so math-heavy and allow for informal thought-experimenting about them. Moreover, the use or non-use of mathematics is not a *necessary* difference between scientific modeling and the performance of STEs because one can reason about models informally and one can perform STEs while using mathematics, even in physics. At most, the difference is a matter of *degree*. As (Frigg, 2010b, p.123–124) wrote:

> Although formalizations play an important role in modeling, not all scientific reasoning is tied to a formal apparatus. In fact, sometimes conclusions are established by solely considering a fictional scenario and without using formal tools at all. If this happens it is common to speak of a thought experiment. Although there does not seem to be a clear distinction between modeling and thought-experimenting in scientific practice, there has been little interaction between the respective philosophical debates. [...] This is lamentable because it seems to be important to understand how models and thought experiments relate to each other. In a recent paper Davies (2007) argues that there are important parallels between fictional narratives and thought experiments, and that exploring these parallels sheds light on many aspects of thought experiments. This take on thought experiments is congenial to the view on models presented in this paper and suggests that modeling and thought experimenting are intrinsically related: *thought experiments (at least in the sciences) are models without formal apparatus.*

I take Frigg's suggestion very seriously. I next introduce Frigg's accounts of scientific models — the fiction view of models — which I shall use to formulate my proposed account of STEs. This account of scientific models, as will become clear in the next Section, was *designed* specifically to account for theoretical models and our imaginative engagement with them. This sounds promising for an account of STEs, to say the least. Moreover, the DEKI-account of representation was designed specifically *for* the fiction view of models. Sartori demonstrated that DEKI applies to STEs straightforwardly. Thus the pieces of the puzzle fall into place.

4.4.2 The fiction view of models

The *fiction view of models* is a peculiar yet up-and-coming contemporary account of scientific models where models are construed as significantly akin to the imagined objects of literary fiction; see notably Godfrey-Smith (2007); Frigg (2010a); Frigg and Nguyen (2016, 2018, 2017a, 2021b, 2020); Levy (2012, 2015); Toon (2012); Weisberg (2013); Salis and Frigg (2020); Salis (2016, 2021, 2020); Levy and Godfrey-Smith (2020). Originally proposed by Godfrey-Smith in 2007, this approach aims to find a middle way between the "over-broad semantic view [of models] and the psychologistic project of Nersessian's mental modelling" (Godfrey-Smith, 2007, p.729). A central aim of the fiction view is to account for the fact that most models in science are treated like and referred to as if they were 'real' things, while most (theoretical) models exist merely in virtue of their description, not as concrete physical objects: "scientists think and talk about model-systems just as ordinary people think and talk about the imaginary characters of fiction." (Salis, 2019, p.9)

The fiction view of models is built on Giere's (1988) account of a model's two-step (indirect) representation of a real-world system: a model description D specifies a model system S, which in turn has relevant similarities with a target system T. Analogously, and taking heed of the lessons learned about representation in the previous Section, proponents of the fiction view of models would say: a model description D describes (and prescribes imaginings about) a model system S, and together D and S represent a target system T (Figure 4.10).[122]

To formulate a full-fledged account of scientific models, which entails notably that one must describe what the *content* of a model is, and that one must describe what *model-based reasoning* and *surrogative reasoning* amount to, the fiction view of models appeals to Walton's theory of fiction (4.7) and his concomitant notion of games of make-believe (4.6). In a

[122] Apparently, Giere strongly opposed using the language of fiction and imagination in philosophy of science, so we should not associate this terminology with Giere. (I thank Roman Frigg for this comment in private communication via e-mail, December 2020.)

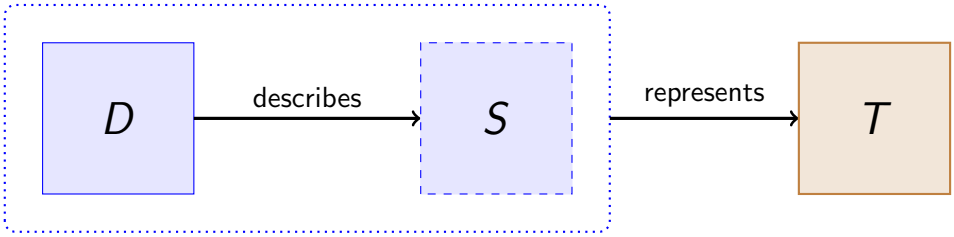

Figure 4.10: Standard schema for the fiction view of models.

nutshell, the idea is as follows.

The *description* of a scientific model is a *work of fiction* in Walton's sense: it is a prop for a game of make-believe. Model descriptions describe the *content* of some model — called the *model system* — and they prescribe imaginings about this content. In terms of Walton's theory of fiction, model systems are *fictional worlds*. The content of model systems is determined by the model's description, which specifies the model's primary fictional truths, together with the model's principles of generation. The principles of generation of a model are the scientific laws, mathematical rules and logical rules of inference, etc. that govern the model (according to the relevant community); they are the *rules* that the fictional world of the model adheres to; they generate the model's indirect fictional truths. See Figure 4.11 for a visualisation.

I again provide some formal definitions for the sake of clarity. According to the fiction view of models, a *model* is an ordered pair

$$M = \langle \mathcal{D}, \mathcal{C} \rangle, \tag{4.15}$$

where $\mathcal{D}$ is the model description, which is a prop for a Waltonian game of make-believe (i.e. it is a work of fiction), and $\mathcal{C}$ is the model content (the topic of $\mathcal{D}$), which is the fictional world of the model description $\mathcal{D}$, and which is generated by the principles of generation (i.e. the laws of nature, mathematical rules, logical rules of inference, etc.) that are associated with that model by the relevant scientific community.

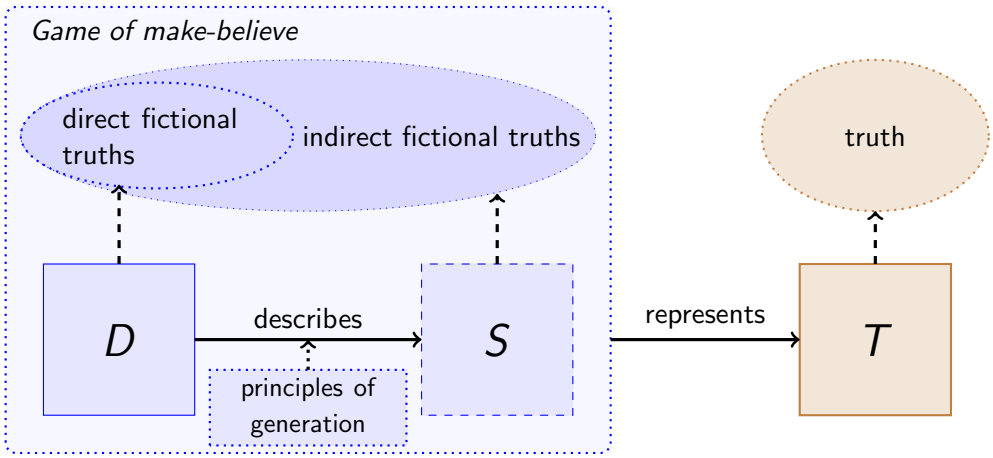

Figure 4.11: Game of make-believe applied to the standard schema for the fiction view of models (Fig. 4.10)

Many models are *representational models*: they are models that represent some target system in the natural world. The fiction view of models straightforwardly accounts for representational models: a model is a representational model iff its description $\mathcal{D}$ denotes some target system T and includes a 'key' that connects the (imagined) entities that make up the (imagined) model and connects them to physical entities pertaining to T, such that the model's i-exemplified properties (4.12) can be translated into properties which can be imputed on T, *à la* the DEKI-account of representation (Section 4.3.4). It bears repeating that *many, but not all,* scientific models are representational models. The fiction view of models can account for both representational models and non-representational models (often called *targetless models*) unproblematically.

Using all this conceptual machinery, the fiction view of models then describes model-based reasoning and surrogative reasoning straightforwardly as follows. To *reason about* a scientific model is to participate in that model's game of make-believe with the aim of discovering, on the basis of the model's direct fictional truths and its principles of generation, the model's indirect fictional truths. Reasoning about scientific models

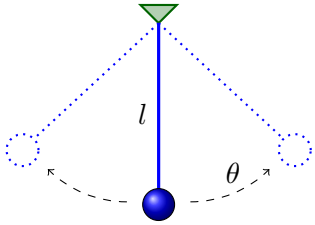

Figure 4.12: Ideal pendulum.

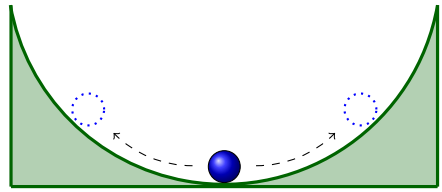

Figure 4.13: Marble in half-pipe.

is therefore a thoroughly *imaginative* affair: we reason about scientific models in our imagination, we reason about models by playing games of make-believe. To *reason with* a representational model about its representandum, i.e. to use representational models for *surrogative reasoning*, is simply to reason *about* a representational model (as described above) *and* to reason about its representation-function with the aim of learning about the model's representandum.

I discuss one familiar example for the sake of illustration.

Consider the model of the ideal pendulum. I can provide the description of this *representational model* with a brief introduction, a non-linear differential equation, and an interpretation that links the variables therein to physical entities and phenomena. This goes as follows.

> **Ideal pendulum:** Imagine a bob hanging on a rigid, massless string, swinging frictionlessly in a plane. See Figure 4.12. The motion of the bob obeys the following differential equation:
>
> $$\frac{d^2\theta(t)}{dt^2} + \frac{g}{l}\sin(\theta(t)) = 0, \tag{4.16}$$
>
> where θ is the angle of displacement away from equilibrium as a function of time, l is the length of the string, and g is the acceleration due to gravity.

Let us reason about this model for a little bit. From its description, we can extract immediately the set of direct fictional truths provided therein. These include: (i) there exists a bob hanging on a rigid, massless string,

swinging frictionlessly in a plane, (ii) the motion of the bob obeys Equation (4.16), and so on. Perhaps the fact that the bob's motion is independent of its mass should also be considered a direct fictional truth. Next, we can also formulate indirect fictional truths using the model's principles of generation. An example of such a principle of generation is the mathematical rule that $\sin\theta \approx \theta$ whenever θ is close to 0. This allows us to formulate the model's indirect fictional truth that, for small angles of displacement away from equilibrium, the motion of the bob approximately obeys the linear differential equation

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{l}\theta(t) = 0, \tag{4.17}$$

which is much easier to solve than Equation (4.16).[123]

Using this example, I next introduce a bit more terminology. Direct fictional truths (i) and (ii) and the indirect fictional truth just mentioned are called *intra-fictional propositions*: they are propositions about the model system made 'while' playing the model's Waltonian game of make-believe. Often in science we *compare* models. In the fiction view of models, this amounts to expressing a proposition that compares i-exemplified properties of *different* fictional worlds. Such propositions are called *inter-fictional propositions*. I can formulate an inter-fictional proposition by introducing another model, e.g. a model of a marble rolling back and forth in a half-pipe (Figure 4.13), and state (in a bit more detail, preferably): the motion of the marble of an ideal pendulum obeys roughly the same equations of motion as a marble rolling back and forth in a half-pipe does. Finally, I can also compare the model of the ideal pendulum with its target system. In this case I express a *meta-fictional proposition*, i.e. a proposition that relates (properties of) the model to (properties of) its target system, on the basis of its representation-relation. I express a meta-fictional proposition if I say e.g. that the motion of the bob of the ideal pendulum accurately represents the motion of the weight hanging on my grandfather's clock.

[123] Solution: $\theta(t) = \theta_0 \cos\left(\sqrt{(g/l)}t\right)$ for initial conditions $\theta(0) = \theta_0$ and $d\theta(0)/dt = 0$.

Salis (2016, p.27) emphasises a crucial difference between intra- and inter-fictional propositions on the one hand, and meta-fictional propositions on the other: only when we express the latter do we exit the model's game of make-believe and assume towards these propositions an attitude of *belief* rather than imagination. Providing a way to evaluate meta-fictional propositions when considering models in terms of games of make-believe is to provide an answer to the following question: how does a scientific model represent the world such that studying a model enables us to learn about the world? Salis (2016) calls this the *problem of model-world comparisons*; Frigg and Nguyen's DEKI-account of representation was designed specifically to deal with this problem, see e.g. (Frigg and Nguyen, 2020, Ch.6–8).[124]

4.4.3 Going forward

Let us now return to Frigg's (2010b, p.123–124) comment that "*thought experiments are models without formal apparatus.*" In the context of the fiction view of models, this comment makes a lot of sense. Thought exper-

[124] Before I continue, I must acknowledge that the fiction view of models did not originate as one account of scientific models for which there is unanimous support, but rather as a *family* of views which differ in some important aspects, particularly pertaining to the question *about what* model descriptions prescribe imaginings: about the (imaginary) model system, or about its *target* system? The divide marks the distinction between *direct* fiction views of models (Toon, 2012; Levy, 2015), which hold that model descriptions prescribe imaginings about the model's *target system*, and *indirect* fiction view(s) of models (Frigg, 2010a; Frigg and Nguyen, 2016, 2018, 2017a, 2021b, 2020; Salis, 2016, 2021), which hold that model descriptions prescribe imaginings, first and foremost, about *model systems*. I chose to adopt the indirect fiction view of models in this Thesis for several reasons: (i) few authors seem to currently endorse the direct fiction view of models, (ii) the indirect fiction view of models seems much more directly applicable to STEs than the direct view, because STEs *are* characterised by imaginative engagement with imaginary scenarios; (iii) the direct fiction view of models cannot account well for non-representational models (or, at least, it becomes indistinguishable from the indirect fiction view); and, most importantly, (iv) the direct fiction view becomes indistinguishable from the indirect fiction view when we *compare* models with the world. That is, is we wish to *evaluate* how, or how accurately, a model represents the world, then we will still end up *comparing some target system with the content of the model, i.e. with the model system.* See e.g. (Frigg and Nguyen, 2020, §6.7) for further criticism of the direct fiction view of models.

iments and models are similar in important respects: the fictional worlds of STEs in Meynell's and Sartori's accounts of STEs and model systems in the fiction view of models, i.e. the fictional worlds of models, are both construed in exactly the same way. Both prescribe imaginative engagement with their content; both are fictional worlds à la Walton. Most importantly: *the fictional worlds of STEs and the fictional worlds of scientific models must be governed by the same type of (scientific) principles of generation* — else they would not be *scientific* TEs or models.

Moreover, the idea that some STEs have an epistemic aim about the world and some have an epistemic aim that just concerns some scientific construct (e.g. a model or theory), coheres perfectly with the idea that some scientific models are representational models and other are not.

I claim that it is, in general, much more straightforward to apply the DEKI-account of representation to *models* than it is to apply them to STEs. As Sartori noted for the case of Maxwell's demon, the fictional worlds of STEs often contain bizarre entities, e.g. demons and the like, that have no place in scientific models, nor in the natural world. These entities may serve to *exemplify* some property of the fictional world of the STE, thus help us with achieving an *internally valid result*, *per* Sartori's distinction (4.14). But these bizarre entities then *distract* us from the STE's representational function, thus make it harder to achieve some *externally valid result* of the STE. In order to figure out how the fictional world of an STE represents some target system in the world (if it does so), and to evaluate if it does so accurately, I suggest we should *reconstruct the representational model that underlies the fictional world of the STE*, i.e. identify the representational model that has the *same* principles of generation as the STE does.

Even STEs that do not represent the world can be analysed with this same method. If an STE does not represent the world, then its epistemic function concerns some *scientific construct*, i.e. a scientific theory, model, hypothesis, concept or what have you. All this scientific thought-experimental reasoning *about* these constructs can be straightforwardly re-

constructed as *reasoning about scientific models*. As Suppes (1960, pp.296-7) suggested: "A Gedanken experiment is given precision and clarity by characterizing a model of the theory which realizes it." The worlds of STEs are governed by models; models *underlie* the worlds of STEs. If this were not the case, then the fictional worlds of STEs are not fictional worlds that are *relevant for science*; if this were not the case, then STEs would be able to teach us very little about scientific constructs.

4.5 The fiction view of STEs

4.5.1 The proposal

The time has come to introduce my proposed account of STEs. To begin, I introduce the core idea of the proposed account:

> **Core idea of the fiction view of STEs:** To perform an STE is to reason (implicitly or explicitly) with and about scientific models with an epistemic aim.

Meynell and Sartori suggested the right type of account of STEs — an account of STEs built explicitly on Walton's theory of fiction — but they did not go far enough. The final step that must be taken before we obtain a full-fledged account of STEs that can handle *both* STEs that only teach us about scientific constructs *and* STEs that teach us about the natural world, is to acknowledge that *scientific models must underlie the fictional worlds of STEs*.

To repeat and make clear, with "underlying" I mean that the fictional worlds of STEs are governed by scientific models: the principles of generation of the fictional worlds of STEs *are* principles of generation of a scientific model. The features of an STE which must remain *invariant* in performances of STEs are facts about a scientific model; the variant features of STEs are *not* part of scientific models. Thus the fictional worlds of STEs can be *reconstructed* as scientific models. Thus, I submit, the descriptions of STEs *include* descriptions of scientific models (in the sense

of the fiction view of models). This inclusion can be explicit, in which case it is unambiguous and evident which model underlies the STE (see examples in next two Sections); or the inclusion can be implicit, in which case we should say that the description of an STE *alludes to* or *implies* the models that govern its world (by introducing some features of the STE which are also facts about a model, and by employing the same principles of generation as that model); see examples in Section 4.2.2.

Some STEs serve to explore *competing models*, and it can even be unclear to the creator of an STE *which* model underlies their *own* STE, as the case of Einstein's photon-box showed us (see more below). But this observation does not imply that it is not always the case that scientific models underlie STEs. Rather the opposite: this observation suggests even more strongly that, in analyzing and evaluating STEs, we should focus our efforts on *discovering* the models that underlie the STE. It is through *this* process that we will understand the fictional world and result of an STE better.

All things considered, the proposed *fiction view of STEs* defines STEs as follows. A *scientific thought experiment* STE is an ordered triple:

$$\mathsf{STE} = \langle \mathcal{D}, \mathcal{F}, \mathcal{E} \rangle, \tag{4.18}$$

consisting of a *description* $\mathcal{D}$, *which implicitly or explicitly includes model descriptions*, a *Waltonian fictional world* $\mathcal{F}$ (4.7), *which partially overlaps with the fictional world of the scientific models described in* $\mathcal{D}$, and an *epistemic aim* $\mathcal{E}$. Thus, what it means to *perform* an STE, according to the proposed fiction view of STEs, is the following:

> Subject S *performs* $\mathsf{STE} = \langle \mathcal{D}, \mathcal{F}, \mathcal{E} \rangle$ iff upon engaging with description $\mathcal{D}$, S plays a game of make-believe (4.6) with $\mathcal{D}$ and reasons about $\mathcal{F}$, with epistemic aim $\mathcal{E}$. (4.19)

I next discuss some immediate consequences of the proposed view.

(i) Theses that make up the fiction view of STEs. To begin, I return to

the five theses that make up Norton's argument view and which Brendel (2018) disentangled (recall Section 4.2.5), and I reformulate these theses such that they make up my proposed fiction view of models (so that Brendel (2018) won't have to do it anymore):

(1) **Identity Thesis.** The invariant features of an STE are facts about a scientific model (or several scientific models); the principles of generation that govern the fictional world of an STE *are* principles of generation of a scientific model (or several scientific models).

(2) **Reconstruction Thesis.** STEs performances can always be reconstructed as instances of *model-based reasoning* based on explicit or tacit assumptions *about scientific models* that yield the same outcome.

  (2a) **Reliability Thesis.** If STEs can be used reliably epistemically, then they must be instances of model-based reasoning (*à la* the fiction view of models) that justify their outcomes or are reconstructible as such instances of model-based reasoning. A thought experiment is a "reliable mode of inquiry" only if the instance of model-based reasoning into which it can be reconstructed justifies its conclusion.

  (2b) **Elimination Thesis.** Any conclusion reached by a (successful) scientific thought experiment will also be demonstrable by a non-thought-experimental form of model-based reasoning.

(3) **Epistemic Thesis.** STEs and the models associated with them have the same epistemic reach and epistemic significance. An STE epistemically justifies its outcome to the same degree as the model described in the STE description justifies its conclusion.

(4) **Empirical Psychological Thesis.** To perform an STE is to reason with and about scientific models.

(5) **Empiricist Thesis**: The result of a thought experiment can only come from experience: "The result of a thought experiment must be the reformulation of [...] experience by a process that preserves truth or its probability." (Norton 2004a, 1142).

These theses, I submit, are all rather plausible. They equate the epistemic scope of STEs to the epistemic scope of *model-based reasoning*. This

provides a much richer view of STEs than e.g. the argument view and the mental-modeling view of STEs did, and it makes much descriptive sense of what scientists *do* when they perform STEs, as I shall argue next.

(i) Relating the fiction view of STEs to the argument view. The argument view of STEs reduced the epistemic scope of STEs to the epistemic scope of arguments. This dismissed notably the epistemic value of imagination in STEs, which, for many, is the main reason to reject the argument view as a full-fledged account of STEs. The proposed fiction view of STEs does not dismiss the epistemic value of imagination: by employing the fiction view of *models*, which construed model-based reasoning explicitly as *imaginative engagement* with fictional worlds, the fiction view of STEs can account for the epistemic value of imagination in STEs *just like* the fiction view of models accounts for the epistemic value of imagination in model-based reasoning. This I take to be a significant improvement over the argument view.

Moreover, I would argue, even if some performance of some STE *is* mostly just argument-based reasoning, then *still* the proposed fiction view of STEs outperforms the argument view. One issue with the argument view of STEs was that it is often not clear *which* argument is presented via STE, and, particularly, how a bizarre, impossible STE-scenario could convey a *sound* argument that can give us true beliefs about the world. The fiction view of STEs provides the answer: if an STE conveys an argument, then it conveys an argument *about a scientific model*; and if this model represents the world, then one can obtain knowledge about the world by performing an STE.

(ii) Relating the fiction view of STEs to the mental-modeling view. Meynell (2014) already explained the benefits of her Waltonian account of STEs over the mental-modeling view: a Waltonian view of STEs can explain the important role of imagination in STEs in a way that does not depend *only* on mental models. Sure, there is often mental modeling going on in the *performance* of STEs. A Waltonian account of STEs can incorporate this unproblematically, as Meynell showed — so too can the

proposed fiction view of STEs. But mental modeling is not *always* relevant in the performance of STEs. Sometimes the performance of STEs is predominantly based on explicit argumenting, in which case the relevant mental state of imagination is *proposition*-imagination, not *action*-imagination. The proposed fiction view of STEs can account for this too, by reconstructing this explicit argumenting as argumenting *about scientific models*, i.e. as model-based reasoning.

(iii) The paradox of thought experiments revisited. The question how we can learn about the world by merely performing a thought experiment had, for decades, a paradoxical air to it; recall Section 4.2.4. The fiction view of STEs dissolves this paradox entirely. There is nothing paradoxical about model-based reasoning. And, since to perform an STE *is*, epistemologically speaking, the same as reasoning with and about models, there is nothing paradoxical about thought-experimenting. By performing STEs, we learn about scientific models; and, if this model represents the world, then, on the basis of this representation-relation, we can justifiably transform our newfound insight about the model into beliefs about the world.

(iv) The heuristic value of STEs revisited. Model-based reasoning has distinct heuristic value, and so does mental-modeling. All this remains true for the fiction view of STEs: the heuristic value of STEs *is* partly the heuristic value of model-based reasoning, partly the heuristic value of mental-modeling, *and*, additionally, partly the heuristic value of fictional narratives. The fiction view of STEs combines all these insights into a coherent whole.

(v) The demonstrative force of STEs revisited. Likewise for the demonstrative force of STEs. If conclusions reached via STEs have *non-inferential justification*, then this justification is reached predominantly through mental modeling and acts of action-imagination, i.e. *quasi-perception*, which is perfectly well accounted for by the fiction view of STEs.

(vi) Demarcating Scientific Thought Experiments. As an added benefit, the proposed fiction view of STEs provides a straightforward *criterion* for the demarcation of scientific thought experiments from non-scientific

thought experiments:

> **Demarcating STEs:** A thought experiment TE is a *scientific* thought experiment iff the invariant aspects of the TE are facts about scientific models, and the principles of generation of TE are principles of generation of those scientific models. (4.20)

(vii) A consistent method for analyzing and evaluating STEs. The proposed fiction view of STEs provides a consistent *method* for analyzing and evaluating STEs: reconstruct the fictional world of an STE by identifying the principles of generation that govern it, and, in doing so, *reconstruct the scientific models that underlie the STE*. Once it is clear which scientific model underlies an STE, we can evaluate whether the STE actually helps us achieve its epistemic aim by evaluating whether we could achieve that epistemic aim by reasoning about the underlying *model*.

(viii) A new light on the epistemology of STEs. I mentioned in the previous Chapter (Section 3.5.1), that knowledge gained through thought experiment may be regarded as a combination of knowledge through imagination *and* knowledge through *testimony*. Thought experiments are deliberately constructed and communicated with the aim of conveying some specific insight, thus this insight is arguably gained partly through testimony (from the one who *constructed* and communicated the (description of the) thought experiment) *and* partly through imagination (by the *performer* of the thought experiment). The proposed fiction view of STEs neatly distinguishes the contributions of these two sources of knowledge, by locating the role of the former (testimony) in the fictional *work* world of the STE and in the construction of the STE description (i.e. the *fictional narrative* of the STE), and by locating the role of the latter (imagination) in the fictional *game* world of the STE and in the way that the performer of the STE *interacts* with the STE description.

I next return to the example STEs introduced in Section 4.2.2 and analyse them using the proposed fiction view of STEs.

4.5.2 Example 1: Galilei's falling bodies

Let us return to Galilei's description of his famous falling bodies STE, which I again quote in its entirety:

> But, even without further experiment, it is possible to prove clearly, be means of a short and conclusive argument, that a heavier body does not move more rapidly than a lighter one provided both bodies are of the same material and in short such as those mentioned by Aristotle. But tell me, Simplicio, whether you admit that each falling body acquires a definite speed fixed by nature, a velocity which cannot be increased or diminished except by the use of force [*violenza*] or resistance. ... If we then take two bodies whose natural speeds are different, it is clear that on uniting the two, the more rapid one will be partly retarded by the slower, and the slower will be somewhat hastened by the swifter. ... But if this is true, and if a large stone moves with a speed of, say, eight while a smaller moves with a speed of four, then when they are united, the system will move with a speed less than eight; but the two stones when tied together make a stone larger than that which before moves with a speed of eight. Hence the heavier body moves with less speed than the lighter; an effect which is contrary to your supposition. Thus you see how, from your assumption that the heavier body moves more rapidly than the lighter one, I infer that the heavier body moves more slowly. (Galilei, 1638, p.)

An analysis per the fiction view of STEs would go along the following lines. Galilei invites the reader of the *Dialogues* to participate in a game of make-believe. He introduces a set of direct fictional truths for an Aristotelian model of falling bodies and generates indirect fictional truths according to certain principles of generation, in order to arrive deductively at a contradictory conclusion. To demonstrate *this* contradiction and to suggest a solution, is the epistemic aim of the thought experiment. I begin by listing the set of direct fictional turths contained in the description from "Each falling body ..." until "... a speed of four."

Direct fictional truths:

(1) Each falling body acquires a definite speed fixed by nature, a speed which cannot be increased or diminished except by the use of force or resistance.

(2) A heavy body falls more rapidly than a light body.

(3) When bodies of different natural speeds are united, the faster body will be retarded by the slower body.

(4) There exists a large stone falling with a speed of eight.

(5) There exists a small stone falling with a speed of four.

(6) The two stones are united into a single body.

That's it, the complete set of direct fictional truths of the game of make-believe that the Aristotelian is happy to participate in. Now, Galilei assists us in formulating several indirect fictional truths, for which we need to appeal only to our basic inferential abilities and two reality-oriented principles of generation. I begin with the principles of generation:

Principles of generation:

(G1) Every body has a mass.

(G2) The mass of a united body is strictly larger than the individual masses of the uniting bodies.

Using these, we can formulate indirect fictional truths as follows.

Indirect fictional truths:

(i) There exists a united system falling with a speed less than eight. (From (1)–(6).)

(ii) The united system is heavier than the large stone falling with a speed of eight. (From (1)–(G2).)

(iii) There exists a united system that is (a) heavier than the large stone falling with a speed of eight, and (b) falling with a speed less than eight. (From (i)–(ii).)

(iv) Contradiction.[125] (From (2) and (iii).)

Thus Galilei arrives at a contradiction when generating indirect fictional truths from the Aristotelian model of falling bodies using basic inferential abilities and reality-oriented principles of generation. Something went wrong. But, contrary to what an argument-like reconstruction would suggest, the STE is not finished just yet; there are still ways for the Aristotelian to remedy the game of make-believe and save its model for falling bodies from contradiction. Gendler (2000b, pp.42–47) mentions several assumptions (principles of generation) that the Aristotelian could introduce to avoid the contradictory result: claim that natural speed or mass is not physically determinate for strapped bodies, for example, or claim that the way the united system behaves depends crucially on the way the two bodies are connected—that it matters whether they should be considered as a single united body or several united bodies.

Gendler goes on to describe that these ways-out are blocked quite naturally by "appeal to broad, defeasible, tacit assumptions, each of which captures an important feature of our representation of experienced reality: natural speed and mass are always physically determined but entification [mereological composition] is *not*" — there is no determinate fact whether strapped-bodies are one object or two. These assumptions are *reality-oriented principles of generation.* They make the Aristotelian conclude that one or more of the (Aristotelian) direct fictional truths — in this case (2), and, consequently, (1) and (3) — are *not* suitable direct fictional truths for a thought experiment about free-falling objects because they lead to contradiction. Instead, the Aristotelian must adopt the Galilean model of free-falling objects, which notably has the direct fictional truth that bodies of different weights fall equally rapidly. With hindsight, we

[125] Contradictions can be fictional truths; we can be *obliged* to imagine a contradiction. (Although it may hard to *actually imagine* them; c.f. Xhignesse (2021).)

know that Galilei's model became an essential part of classical mechanics, thus we can genuinely say that the Aristotelian gained new scientific knowledge by adopting the Galilean model of free-falling objects.

The argument view and the mental-modeling view of STEs can both account for Galilei's falling bodies only partly. Performing Galilei's falling bodies surely *involves* argumenting and it also plausibly *involves* mental-modeling. But these activities do not point at the *epistemic aim* of the STE: the epistemic aim of this STE is to argue — by means of a combination of an illustrative argument and some instructions for mental-modeling (notably pertaining to "blocking the ways out", see above) — *in favor of one scientific (representational) model over another*. The fiction view of STEs can account for this in ways that go beyond the argument view and the mental-modeling view. By redirecting the focus of these respective towards the *scientific models* underlying STEs, my proposed fiction view of STEs incorporates the strengths of the argument view and the mental-modeling view in one coherent conceptual framework.

4.5.3 Example 2: Clement's Sisyphus

To repeat: Clement (2009a, 2018), in researching the role of analogy and imagery in scientific reasoning, presented several test-subjects (not physicists) with diagram 1A of Figure 4.2 and the following description (Clement, 2009a):

> You are given the task of rolling a heavy wheel up a hill. Does it take more, less, or the same amount of force to roll the wheel when you push at X, rather than at Y? Assume that you apply a force parallel to the slope at one of the two points shown, and that there are no problems with positioning or gripping the wheel. Assume that the wheel can be rolled without slipping by pushing it at either point.

Diagrams 1B and 1C in Figure 4.2 represent the analogies that one test-subject, call her Alice, made in finding the solution to this thought experiment: first, Alice imagined a physical system with which she has real-life

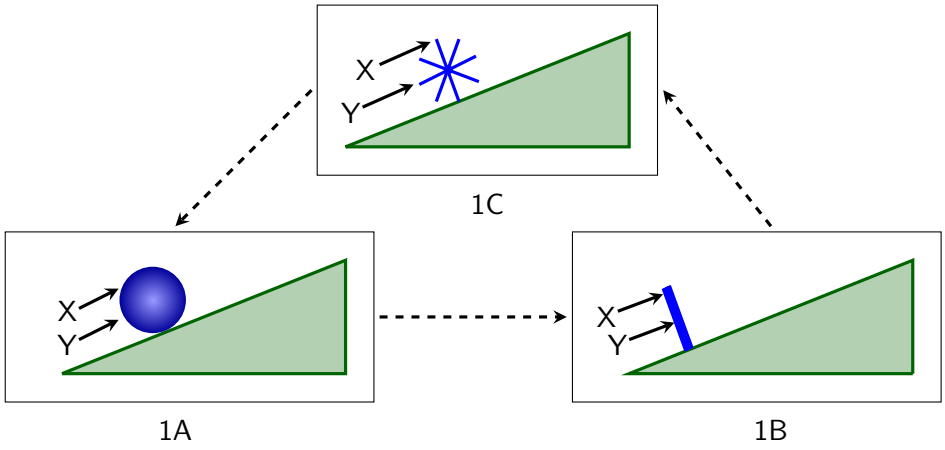

Figure 4.14: Clement's Sisyphus. Based on (Clement, 2009a, Figure 1), original description: "Analogies for Sisyphus problem."

experience— lever-like system 1D —and knew the answer in this context.[126] Then, she convinced herself that you can make up a wheel by superimposing many levers (1C), thus justifying the intuition that her answer to 1B carries over to 1A.

Clement presents a coherent analysis of the analogical reasoning occurring in this STE: he describes how Alice searched first for an analogy (a lever), then for a 'confident base-case for the analogy' (the lever-like system 1B), and finally for 'bridging cases' (the rimless wheel 1C) to ground confidence in the original analogy and in being able to transfer the conclusions drawn therein onto the original situation 1A. The only downside that I see with this analysis is that it employs the vocabulary of cognitive psychology and (elsewhere) Nersessian's mental modeling account of model-based reasoning. I do not pretend to improve on the content of his analysis. I only claim that the fiction view of models *also* can make sense of this thought experiment, but now with a set of concepts perhaps more

[126] Other test-subjects make different analogies, see Clement (2009a). This can be accounted for unproblematically by the fiction view of models: different agents initiate different game of make-believe, thus formulating different inter-fictional propositions, but the same analysis applies.

adequate for a universal account of STEs.

According to the fiction view of STEs, the analogies drawn by Alice are expressed by *inter-fictional propositions*. Diagrams 1A, 1B and 1C are descriptions of three different models: they are props in three different games of make-believe, each with their own direct fictional truths. In the game of make-believe of 1A, the original task, Alice lacks the principles of generation to generate any of the three possible indirect fictional truths. This is the epistemic aim of the STE: given game 1A, find a way to generate any of the three possible indirect fictional truths (does it take more, less, or the same amount of force to push the wheel at X, rather than at Y?). Alice's search for the answer can be reconstructed as follows:

Game 1A (unfinished)
Direct fictional truths:

(A.1) There exists a heavy wheel.

(A.2) The wheel is rolled up a hill.

(A.3) The wheel can be pushed unproblematically at X or at Y.

(A.4) It takes force to roll the wheel up the hill.

Evidently, Alice lacks principles of generation needed to generate either of the following three candidate indirect fictional truths: (A.i) It takes less force to roll the wheel at X rather than at Y. (A.ii) It takes more force to roll the wheel at X rather than at Y. (A.iii) It takes the same amount of force to roll the wheel at X rather or Y. In her search for a principle of generation that can lead her to any of the indirect fictional truths, she initiates a new game of make-believe concerning a model *possibly* analogous to Game 1A. In game 1B, two principles of generation can be reconstructed. The first is provided by Alice's experience with real-world lever-like systems, and is presumably formulated by Alice through mental-modeling: *it is easier to push over the lever at X than at Y*. The second is a reconstructed, implicit principle used to connect such experience to

the theoretical concept 'force', necessary for a valid argument-style reconstruction of this STE: *If something is easier to push over, then it takes less force to push over.* (This principle shows itself in Alice's *mental-modeling* efforts.) Together, these principles allow Alice to formulate an indirect fictional truth of 1B.

Game 1B
Direct fictional truths:

(B.1) There exists a lever.

(B.2) The lever can be pushed over unproblematically at X or at Y.

(B.3) It takes force to push over the lever both at X and at Y.

Principles of generation:

(B.G1) It is easier to push over the lever at X than at Y.

(B.G2) If something is easier to push over, then it takes less force to push over.

Indirect fictional truth:

(B.i) It takes less force to push over the lever at X than at Y. (From (B.1)–(B.G2).)

Now, Alice initiates the third game, which establishes her confidence in the analogy between games 1A and 1B, and, consequently, enables her to formulate an inter-fictional proposition between games 1A and 1B. Note that here the indirect fictional truth from Game 1B is now used as a principle of generation in Game 1C.

Game 1C
Direct fictional truths:

(C.1) There exists a rimless wheel, consisting of superimposed levers.

(C.2) The rimless wheel is pushed up a hill.

(C.3) The rimless wheel can be pushed unproblematically at X or at Y.

(C.4) It takes force to push the rimless wheel up a hill.

Principle of generation (from 1B*):*

(C.G1) It takes less force to push over a lever at X than at Y.

Inter-fictional proposition (between 1B *and* 1C*):*

(C.IF) If it takes less force to push over a lever at X than at Y, then it takes less force to push a rimless wheel up a hill at X than at Y.

Indirect fictional truth:

(C.i) It takes less force to push a rimless wheel up the hill at X than at Y. (From (C.G1)–(C.IF).)

Finally, Alice returns to game 1A. She now imports the implied f-truth (C.i) into the principles of generation for game 1A, and formulates an inter-fictional proposition between games 1C and 1A that enables her to use (C.i) to generate the desired implied f-truth in game 1A.

Game 1A (finished)

Primary fictional truths:

(A.1) There exists a rimmed wheel.

(A.2) The wheel is rolled up a hill.

(A.3) The wheel can be pushed unproblematically at X or at Y.

(A.4) It takes force to roll the wheel up the hill.

Principle of generation (from 1C*):*

(A.G1) It takes less force to push a rimless wheel up a hill at X than at Y.

Inter-fictional proposition (between 1C *and* 1A*):*

(A.IF) If it takes less force to push a rimless wheel up a hill at X than at Y, then it takes less force to push a rimmed wheel up a hill at X than at Y.

Indirect fictional truth:

(A.i) It takes less force to push the rimmed wheel up the hill at X than at Y. (From (A.1)–(A.IF).)

Alice has now solved the original problem of formulating one of three possible indirect fictional truths for diagram 1A. Her epistemic aim is achieved, the STE is completed. Again, Alice never left the realm of fiction — she never stopped playing games of make-believe. She has only learned about models.

In this reconstruction of Alice's performance of Clement's Sisyphus according to the fiction view of models, the insights of both the argument-view (there are arguments occurring in Alice's performance of the STE; indeed, her entire performance can be reconstructed as a series of arguments) and the mental-modeling view (there is mental-modeling occurring in Alice's performance of the STE, which played a crucial epistemic role for obtaining results and establishing analogy-relations) are wholly incorporated. I consider it an advantage for my proposed account that it can do so in a single, coherent conceptual framework.

Lastly, I note that different test-subjects who performed this same STE made different analogies (Clement, 2009a). This is unproblematically accounted for by the proposed fiction view. Different analogies only means that there are different models compared with each other — but the description of the STE and its epistemic aim remain the same. By reconstructing the scientific models that underlie these analogies, we are even in a position to *compare* the quality of the analogies by investigating the representation-relation between the models. If an analysis of these others performances of this same STE were to proceed along the lines of the argument view or the mental-modeling view of STEs, then one must employ an account of scientific models alongside their respective account of STEs. Because the proposed fiction view of STEs is explicitly built on the fiction view of *models*, the fiction view of STEs already possesses all the necessary conceptual tools for this task.

4.5.4 Other examples revisited

I next briefly review the other four example STEs (Section 4.2.2) in light of the proposed fiction view of models.

(a) Newton's bucket. With his bucket thought experiment, Newton aimed to convey an argument against the Cartesian idea that motion can be defined as relative motion of an object with respect to its nearest neighbouring object. When I introduced Newton's bucket, I emphasised in particular that this STE presents a physically impossible scenario (a bucket spinning in otherwise empty space) and that the description of this STE left implicit the crucial assumption that the bucket is spinning in otherwise empty space, which was notably objected to by Ernst Mach.

In light of the fiction view of models, the description of Newton's bucket presents a model of a spinning bucket of water and specified the following epistemic aim: how ought we to define motion in this model? Newton aims to demonstrate that Descartes' definition does not work for this particular model, thus demonstrating that his own alternative definition is better.

But the STE description evidently underdetermined its fictional world, i.e. the description of Newton's bucket did not adequately specify the principles of generation for the model system that underlies the STE. As Mach's objection to Newton's bucket showed us, there is plenty room for disagreement about *which* particular model Newton presented: a model of a bucket in otherwise-empty space, or a model of a bucket in a universe filled with matter? The difference is important because it has consequences for reaching the STE's epistemic aim. The fiction view of STEs points us to this difference.

(b) Maxwell's demon. With his demon thought experiment, Maxwell aimed to illustrate the statistical nature of the Second Law of thermodynamics. To do so, Maxwell presented a relatively simple (classical) *model* of molecules in a vessel divided by a diaphragm. He then *added* to this model a remarkable, hypothetical entity that served to illustrate the statistical nature of the Second Law by presenting a scenario where the Second

Law of thermodynamics is violated. The epistemic aim of the STE can be interpreted as the question: is the demon physically realizable, in the sense that one can obtain the same result (a violation of the Second Law) without help from a demon?

As I said before, Maxwell's demon has gotten quite a life of its own, as in the past century it has been investigated whether the demon is physically realizable from the perspective of various, *different* scientific theories — in the context of various *models* of these various theories. The fiction view of STEs tells us what is going on here: the principles of generation of the original STE change, thus the models that *underlie* this STE change. Strictly speaking, then, we should probably say that the STE itself has changed its identity — that there are *more than one* distinct versions of Maxwell's demon: one in classical mechanics, one in quantum theory, one in information theory, etc. To exhaustively analyse the differences between these views would be an expansive project; see e.g. Myrvold (2011) for an important, historically nuanced starting point. Importantly, one would analyse the differences between these views by reconstructing the *models* that underlie these distinct versions of Maxwell's demon. All this vindicates, in my view, the fiction view of STEs.

Before I move on, I wish to note only one more thing about Maxwell's demon, which seems to have been curiously overlooked in discussions of it. Namely, that Lord Kelvin (Thomson, 1874), who was responsible for giving the name "demon" to Maxwell's hypothetical thought-experimental entity, *already demonstrated the physical realizability of the 'demon' (i.e. he demonstrated a violation of the Second Law) by introducing a thought experiment about reversing time in an explicit (classical statistical) scientific model of a gas of molecules that does not include a "demon"*, immediately after he first introduced Maxwell's "demon". Remarkably, in doing so, Lord Kelvin even anticipated Poincaré's recurrence theorem[127] in the process. He wrote (Thomson, 1874, pp.442-3):

[127] Roughly speaking, Poincaré's recurrence theorem states that dynamical systems such as ideal gases in a closed environment will, after sufficiently long (finite) time, return to their original state.

If no selective influence, such as that of the ideal "demon," guides individual molecules, the average result of their free motions and collisions must be to equalise the distribution of energy among them in the gross; and after a sufficiently long time from the supposed initial arrangement the difference of energy in any two equal volumes, each containing a very great number of molecules, must bear a very small proportion to the whole amount in either; or, more strictly speaking, the probability of the difference of energy exceeding any stated finite proportion of the whole energy in either is very small. Suppose now the temperature to have become thus very approximately equalised at a certain time from the beginning, and let the motion of every particle become instantaneously reversed. Each molecule will retrace its former path, and at the end of a second interval of time, equal to the former, every molecule will be in the same position, and moving with the same velocity, as at the beginning; so that the given initial unequal distribution of temperature will again be found with only the difference that each particle is moving in the direction reverse to that of its initial motion. This difference will not prevent an instantaneous subsequent commencement of equalisation, which, with entirely different paths for the individual molecules, will go on in the average according to the same law as that which took place immediately after the system was first left to itself.

By merely looking on crowds of molecules, and reckoning their energy in the gross, we could not discover that in the very special case we have just considered the progress was towards a succession of states in which the distribution of energy deviates more and more from uniformity up to a certain time. The number of molecules being finite, it is clear that small finite deviations from absolute precision in the reversal we have supposed would not obviate the resulting disequalisation of the distribution of energy. But the greater the number of molecules, the shorter will be the time during which the disequalising will continue; and it is only when we regard the number of molecules as practically infinite that we can regard spontaneous disequalisation as practically impossible. And, in point of fact, if any finite number of perfectly elastic molecules, however great, be given in motion in the interior of a perfectly rigid vessel, and be left for

> a sufficiently long time undisturbed except by mutual impacts and collisions against the sides of the containing vessel, it must happen over and over again that (for example) something more than nine-tenths of the whole energy shall be in one half of the vessel, and less than one-tenth of the whole energy in the other half.

Moral of the story: if only we had paid more attention to the *model* that underlies the original version of Maxwell's demon, rather than its distracting "demon", then perhaps it would not have been necessary to create all the subsequent variations of this STE. The *external validity* (Sartori, 2023) of Maxwell's demon was evident — and even *explicitly mentioned* — from the very moment that the STE was introduced.

(c) Einstein's photon-box. With his photon-box thought experiment, Einstein intended to provide a counter-example to Heisenberg's uncertainty relation. But, as Bohr's response to this STE showed us, it turned out that *Einstein was mistaken about the models underlying his own STE.* Einstein's version of the photon-box appealed only to quantum mechanical principles of generation. Bohr responded by demonstrating that *general-relativistic* principles of generation were required for an adequate analysis of the thought-experimental scenario (the photon-box). Interestingly, although Bohr's response allegedly shocked (and partially convinced) Einstein, ambiguity remains to this day about the *correct* analysis of this thought-experiment; see e.g. (Beller, 1999; de la Torre et al., 1999; Marage and Wallenborn, 1999; Hilgevoord, 2002; Howard, 2007; Schmidt, 2022).

From the perspective of the fiction view of STEs, this episode from the history of science shows us that (i) the creator of an STE can be mistaken about the models that underlie their own STE, and even (ii) that ambiguity can *persist* about which models underlie an STE. STE's are objects of investigation; we can *discover* things about STEs. What we can discover about STEs, is *which models underlie them.*

(d) Norton's dome. With his dome thought-experiment, Norton (2008) aimed to demonstrate that the theory of Newtonian mechanics is not deterministic. He did so by introducing a Newtonian *model* of a dome of

a particular shape and demonstrated that it allows of 'indeterministic' solutions.

Norton's dome is an example of an STE that does not have an epistemic aim concerning the natural world. This STE directly concerns a model; there is no ambiguity about *which* model underlies the STE. Notably, this model (the ball on the dome) is a *targetless, i.e. non-representational model*: the model does not represent some target system in the world. Norton's dome is a prime example of a thought experiment that is an *explicit* form of model-based reasoning: the description of this STE explicitly *includes* the explicit description of a model.

The fiction view of STEs can account for this unproblematically. It does not matter for our analysis whether an STE concerns a representational model or a non-representational model. All STEs can be analysed by one and the same method: reconstruct the scientific models that underlie the STE, and analyse *those*.

4.5.5 Further results

The conceptual frameworks that I introduced throughout this Chapter, on which the proposed fiction view of STEs is built, are quite elaborate: Walton's theory of fiction, Waltonian games of make-believe, and the fiction view of models, all of these conceptual frameworks contain many concepts that do not seem *directly* relevant for understanding STEs — although I have argued that most of them *are* relevant. But one may wonder: is there really much benefit to employing such an extensive conceptual framework to account for 'just' thought experiments?

I think there is. So far, I have discussed the most prevalent and most directly pressing questions concerning STEs that have been explicitly accounted for by well-known accounts of STEs such as the argument view and the mental-modeling view. But there is a long list of more niche questions, of e.g. ontological, metaphysical, and even sociological character, pertaining to STEs. I next argue that the fiction view of STEs provides answers to all these questions too.

(a) Question: if an STEs has a logical structure, what modifications to this structure does it "tolerate before it ceases to exist and a new one is born?" (Brown and Fehige, 2019, p.1). As Meynell said: fictional worlds of STEs are what confer identity. So, bracketing those STEs where the ambiguity about the fictional world *is* the epistemic aim of the STE (e.g. STEs that are intended to explore two conflicting models), an STE 'ceases to exist' when its fictional world changed too much. How much? Simple: when a *different* model underlies the STE.

(b) Question: what variations do performances of TEs allow while remaining (a performance of) the *same* STE? (Kujundzic, 1998). Simple. Performances of STEs are reconstructed as Waltonian *game worlds* (4.9). When we participate in a game of make-believe, we have the *intention* that our game worlds overlap with the fictional *work worlds* (4.8) of the STE. Thus we have an answer to the above-mentioned question: we perform the *same* STE as long as we *intend to overlap our game world with the work world same STE*. So, to *which* STE any given STE-performance pertains is wholly dependent on *intention*. If there is plenty overlap, but there is no intention, then the STE is not performed. If there is little or no overlap, but there is all the intention, then the STEs *is* performed, but it is just incredibly badly performed. If there is overlap *and* intention, then the STE is performed *well*.

(c) Question: what happens when an STE admits of different or even contradictory interpretations — so-called "TE-anti-TE pairs"? Are we then concerned with one STE or multiple STEs? (Norton, 2004b; Meynell, 2014). In general, we are concerned with one STE. If the epistemic aim *is* to figure out which model underlies the fictional world of the STE, then we are evidently concerned with one STE. Same story if there is agreement about which models applies but there is disagreement about the indirect fictional truths of the model — or about its direct fictional truths, such as Mach's response to Newton's bucket told us.

(d) Question: What about the social aspect of STEs: must all the epistemic value and 'evidential significance' be found 'in' the STE or is the

social context in which the STE was created important to acknowledge? (McAllister, 1996; Potters and Leuridan, 2004). The fiction view of STEs highlights the *social* aspect of STEs. The fiction view of STEs ascribes inter-subjectively stable existence to STEs, in a sense that goes beyond mental states. But it does not ascribe wholly *objective* existence to STEs. STEs are works of fiction, and works of fiction are always works of fiction *relative* to a community. This community and its shared conventions and communal background beliefs are crucial to generating an STEs fictional world. Thus there is a crucial *social* dimension to thought-experimenting. This dimension is not at all captured by the argument view or the mental-modeling view. But it is captured straightforwardly by the fiction view of STEs: Waltonian games of make-believe are held together by social conventions — scientific models are held together by *scientific* (social) conventions.

By capturing the social aspect of STEs, the fiction view of STEs enables us to reply to e.g. McAllister (1996) and Potters and Leuridan (2004), who object that the "social aspect" of a thought experiment is not explained in existing accounts of STEs: the "evidential significance" of an STE lies not wholly in the STE itself, they argue, it lies at least partly in the thought-experimenter and its interaction with the scientific community. *Which* models are relevant for a particular STE depends strongly on social factors: the scientific community determines which models are currently accepted as 'valid' scientific models. All this is accounted for by a Waltonian account of STEs, so too the fiction view of STEs, as Salis and Frigg (2020, p.44) note:

> [Waltonian] make-believe has an objective content that is normatively characterised in terms of social conventions implicitly or explicitly understood as being in force within the relevant game. The social character and objectivity of make-believe are typical for the sort of imaginative activities involved in TEs and [scientific] modeling.

(e) Question: why do STEs figure more prominently in e.g. physics

and economics but less so in biology and chemistry (Brown, 1991; Stuart, 2019)? Presumably, part of the answer to this question resides in the role of models in the respective field: physics and economics are scientific disciplines that are *centered* around the practice of formal modeling, biology and chemistry arguably less so. I argued that thought-experiment *is* reasoning about models. It should be no surprise, then, that the practice of thought-experimenting is more ingrained in formal-model-heavy scientific disciplines than in other scientific disciplines.

(f) Question: what ontological commitments does my proposed account demand (El Skaf and Stuart, 2023)? Admittedly, I do not consider this a very interesting question. But it is useful to see, at least in sketch, how the proposed fiction view of STEs outperforms the argument view and the mental-modeling view. As a result, however, most systematic accounts of STEs do not directly provide answers to these questions nor do they employ the conceptual tools that enable us to formulate answers straightforwardly. I mention one example discussed by El Skaf and Stuart (2023, §2.2): thought experiments are often identified with *arguments*, but it is not often explicitly discussed *what arguments are*. It can be argued that arguments are made up of propositions, inferences or instructions for inferences, or what have you, and each of these positions entail different ontological and metaphysical commitments that are rarely discussed explicitly for the case of STEs. The proposed fiction view of STEs is wholly explicit about its ontological commitments, as should be evident from explications (4.18) and (4.19).

Of course, there are plenty metaphysical issues haunting fiction. Many objections have been raised against Walton's theory of fiction and to the fiction view of models. The fiction view of STEs directly inherits these problems. But these problems have been responded to many times, in my eyes satisfactorily; see notably Walton (1990) and Frigg and Nguyen (2021b) and references therein, but see also Kripke (2013). I do not regard the proposed fiction view of STEs to raise new metaphysical issues beyond these familiar issues pertaining to the metaphysics of fiction.

(g) Question: we often use thought experiments in teaching — how can my proposed account of STEs improve the *quality* of the way we teach with thought experiments (Reiner, 1998; Reiner and Burko, 2003; Reiner and Gilbert, 2000; Steier and Kersting, 2019; Kersting et al., 2021)? My proposed account of STEs can help us improve the way we teach with thought experiments in three notable ways, corresponding to the three conceptual frameworks that it is built upon: (i) Walton's theory of fiction, (ii) Walton's notion of games of make-believe, and (iii) the fiction view of models.[128] I briefly discuss each in turn.

(i) Because Walton's theory of fiction emphasises the role of *works of fiction* as *props* for our engagement with fiction, the fiction view of STEs draws attention to the pedagogical importance of a good STE *description*. An STE description describes the fictional world of that STE, and it prescribes imaginings about it. A *good* STE description describes the fictional world *well*, in the sense that it draws imaginative attention to the features of this scenario that will help us in achieving the STE's epistemic aim. This is a lesson for those who wish to teach with thought experiments: pay attention to the way you *present* them. The details of the description matter a lot.

(ii) Because Waltonian games of make-believe are fundamentally *social* activities, the fiction view of STEs draws attention to the social aspect of STEs too. When students perform thought experiments in the classroom, they often perform them *together*. The conceptual framework of Waltonian games of make-believe enables us to capture what happens here; recall my discussion at question *(g)* above.

(iii) Most importantly, the fiction view of STEs is built on the fiction view of models: the core assumption of the proposed view is that *scientific model underlie STEs*. This gives us explicit direction for how to *construct* STEs efficiently: focus on creating an STE — and an STE description — that unambiguously exemplifies features of scientific models that are relevant for the pedagogical purpose at hand. This again harks back to

[128] I presented these suggestions at a conference; Rijken (2021c).

Suppes' (1960, pp.296-7) comment that:

> A Gedanken experiment is given precision and clarity by characterizing a model of the theory which realizes it.

I submit that the converse is also true: a scientific model is given precision and clarity by performing an STE that exemplifies that model. If we teach science, we teach predominantly about scientific models — if we teach about models, we can, and should, use scientific thought experiments.

4.6 Conclusion

To conclude and recapitulate, in this Chapter, I have proposed a novel account of STEs that is explicitly built on the recently-developed fiction view of models, and which improves on recent proposals by Meynell (2014) and Sartori (2023).

I began by discussing the two 'core questions' concerning STEs: (I) what are STEs, and (II) what, and how, do we learn by performing STEs (Section 4.2.1). I then introduced two example STEs at length (Galilei's falling bodies and Clement's Sisyphus) and briefly introduced four more STEs, each of which exhibited different, characteristic features of STEs (Section 4.2.2).

With these examples in hand, I then returned to core question (I) and defined what STEs are (4.3) and explicated what it means to perform an STE (4.4) (Section 4.2.3). In the subsequent Section (4.2.4), I returned to core question (II) and discussed in particular: (i) that STEs are considered to have distinct *heuristic value* and *demonstrative force*, both of which must be explained by an account of STEs, (ii) the 'paradox of thought experiments', which is much less paradoxical than it long seemed to be, and (iii) the difference between STE-beliefs and quasi-perceptual beliefs, which were the topic of the previous Chapter. I then evaluated two long-standing accounts of STEs — the argument view (Section 4.2.5) and the

mental-modeling view (Section 4.2.5) — noting their respective strengths and weaknesses.

I then discussed the relation between STEs and the concept of *fiction*. I first motivated the idea that the concept of fiction is useful for understanding STEs (Section 4.3.1). I then introduced and discussed at length the theory of fiction from Walton (1990), and I provided explications for the core concepts in this theory of fiction (4.7) and its underlying conceptual schema of 'games of make-believe' (4.6). I then discussed the account of STEs proposed by Meynell (2014), which is explicitly built on Walton's theory of fiction: Meynell reconstruct the imaginary scenario of STEs as *fictional worlds à la* Walton. Meynell's account of STEs was a step in the right direction, but it was (by Meynell's own admission) only a *partial* account of STEs. What was missing, was a consistent method for evaluating how the fictional world of an STE *relates* to the natural world, and, consequently, how the fictional world of the STE provides evidence for, or *justifies*, insight gained about the natural world by performing the STE. I then discussed the account of STEs proposed by Sartori (2023), who promised to improve on Meynell's account by introducing an explicit account of scientific representation: DEKI-representation. I concluded that Sartori succeeded only partially in improving on Meynell's account (Section 4.3.4), notably because Sartori's account *still* does not provide a consistent method for reconstructing the fictional world of an STE and for evaluating the relation between this fictional world and the natural world.

My suggestion: to complete Meynell's and Sartori's account of STEs, we should reconstruct the fictional world of an STE as a *scientific model*. To motivate this suggestion, I first discussed the close relation between STEs and scientific models, which has been noted many times before, even by Suppes (1960), but which nonetheless has been relatively underexplored in the literature on STEs (Section 4.4.1). I then introduced a recently-developed account of scientific models that is particularly well-suited for formulating a full-fledged Waltonian account of STEs: the fiction view of models (4.15).

Finally, with all these pieces of the conceptual puzzle in hand — Walton's theory of fiction (4.7), Walton's notion of games of make-believe (4.6) and the fiction view of models (4.15) — I proposed my own account of STEs: the fiction view of STEs (4.18). According to the proposed account, to perform a scientific thought experiment is to reason with and about scientific models (4.19). I formulated the proposed view as clearly as possible and noted immediate consequences of the proposed view (Section 4.5). I illustrated how STEs are analysed in light of the proposed fiction view of STEs by analysing at length Galilei's falling bodies (Section 4.5.2) and Clement's Sisyphus (Section 4.5.3). I also commented on noteworthy aspects of the other four example STEs (Section 4.5.4), thereby demonstrating the fruitfulness of the proposed account. Finally, I discussed how the proposed fiction view of STEs also provides answers to niche questions about STEs that have been posed occasionally in the literature but which have been under-illuminated, if not totally ignored, by other accounts of STEs.

Chapter 5

Conclusions

En hoe verder hij ging, des te langer was zijn terugweg.

C.C.S. Crone, *Het Feestelijk Leven*

5.1 Summary of results

This Thesis draws to a close. In this Concluding Chapter, I summarise the results of each Chapter and I indicate directions for future research.

5.1.1 Chapter 2: Explicating Imagination

In the first main Chapter, *Explicating Imagination*, I dealt with the following three research questions:

- ➢ How can we explicate the mental state of imagination?
- ➢ What are core characteristics of the mental state of imagination?
- ➢ How does imagination relate to similar mental states, notably to perception, belief, visualisation, supposition and memory?

I began by noting the Divide between *Imagers*, who believe that mental states of imagination necessarily have sensory content (2.4), and *Wide-*

heads, who believe that mental states of imagination do not necessarily have sensory content (2.5) (Section 2.3.1). I next distinguished two *types* of mental states of imagination: proposition-imagination and action-imagination (Section 2.3.2). I argued that a third oft-mentioned type of imagination, *entity-imagination*, is not a distinct type of imagination but rather reduces to either proposition-imagination or action-imagination (2.9): imagining an entity ε is either imagining *that* ε exists (proposition-imagination) or imagining *seeing* ε (action-imagination) (Section 2.3.3).

I then turned to *explicating* proposition-imagination, which I explicated as follows (Section 2.4.1):

[ImProp] S imagines that p iff S occurrently accepts that p is τ-possible, for some appropriate modality type τ. (2.13)

I argued that this explication captures eight characteristics which have been mentioned by many authors as *typical features of imagination*, and also that it highlights in which cases imagination does *not* exhibit these features (Section 2.4.2):

I. imagination is *episodic* and *temporary*,

II. imagination is *voluntary* and *deliberate*, at least more so than e.g. beliefs and hallucinations — although I flagged many times that the subtle notion of 'voluntariness' must be handled with extreme care,

III. imagination involves thinking of things as *possible*,

IV. imagination is *logically independent* of perception and belief,

V. imagination is largely *quarantined* from physical action and other typical direct consequences of perceptions and beliefs — although I noted that imagination often *indirectly* motivates physical action,

VI. imagination is '*belief-like*', notably in the sense that the inferences we make with imagined propositions "mirror" inferences that we would make if we believed those propositions,

VII. imagination is '*perception-like*', in the sense that 'imagistic' types of mental states of imagination have sensory content, just like perception does, and

VIII. an imagined proposition *under-determines* the content of a mental state of proposition-imagination, notably to the extent that a choice of *topic* under-determines the imagined *content*: content of a mental state of imagination is under-determined by "top-down" choices.

With this explication of proposition-imagination (2.13) in hand, I then explicated *supposition* (2.15), *counterfactual thought* (2.16), and *conceiving* (2.21) as distinct *types* of proposition-imagination that have an epistemic purpose — and, additionally, having the counterfactual thought that p (2.16) implies disbelieving that p (Section's 2.5.1–2.5.3).

I explicated proposition-*visualisation* (2.25) and proposition-*picturing* (2.28) as mental states of proposition-imagination that additionally have *representing*, and *accurately representing*, sensory content in their explications, respectively. Analogously, I explicated *entity*-visualisation and *entity*-picturing of some entity ε as visualising (2.26) and picturing (2.29), respectively, the proposition *that* the entity ε exists (Sections 2.5.4–2.5.5).

I then turned to explicating action-imagination (Section 2.6). I distinguished two types of action-imagination, action-imagination 'from the inside' and 'from the outside', which I explicated as follows (Section 2.6.1):

[ImActIn] Subject S *inside-imagines* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate type of modality, and S has an occurrent endogeneous mental state such that its sensory content represents the event of S ϕ-ing.

[ImActOut] Subject S *outside-imagines* ϕ*-ing* iff S accepts that it is possible that S is ϕ-ing, for some appropriate type of modality, and S has an occurrent endogeneous mental state such that its motor content represents the event of S ϕ-ing. (2.30)

I emphasised in particular that imagining an action (2.30) implies *dispositionally* accepting that the action is possible — this is the case because (I argued that) imagining implies accepting a possibility. If you also *occurrently* accept this possibility, then you also imagine the proposition *that* you perform the action. Such is the logical connection between proposition-imagination (2.14) and action-imagination (2.30).

With the explications for the two types of action-imagination (2.30) in hand, I then revisited what it means to visualise (2.32) and picture (2.33) *actions* — both of these mental states were explicated straightforwardly on the basis of the preceding results (Section 2.6.2).

Finally, I turned to the relation between imagination and *memory* (Section 2.6.3). I began by explicating a mental state of *mnemonic imagination* (2.35), i.e. a mental state of action-imagination (2.30) with mnemonic content. I then argued that so-called *episodic imagination* (2.37), i.e. a mental state of memory with sensory content (of past perceptions), *is* mnemonic imagination (2.35). Lastly, I turned to so-called *proposition-memory* (2.41), i.e. a mental state of memory with propositional content (of past knowledge), which, somewhat surprisingly, turned out to be *un*related to imagination.

I presented a schematic overview of inter-relations between the concepts explicated in this Chapter — a *conceptual geography of imagination and allied concepts* — in Figure 2.1 (Section 2.2, p.22).

I concluded this Chapter by elaborating on a phenomenon that imagination often exhibits in practice: having a mental state of one type of imagination typically comes "accompanied" by other types of imagination (Section 2.8.1). I then turned to the cognitive-scientific perspective on imagination, noting two typical regularities pertaining to this notion of "accompaniment" that have been empirically researched in recent years: (i) the fact that, with 'imagistic' types of imagination, our eyes typically move just like they would if the mental state were a mental state of perception, and (ii) the fact that imagining an action from the inside (2.30) causes your neural sensory processing mechanisms to *anticipate* and *pre-*

pare for sensory input that would come if the imagined action were actually performed. I noted that these regularities are interesting, but that further empirical results are required before these *typical* regularities bear consequences for *explications* of imagination.

All things considered, I believe that I have provided elaborate and conclusive answers to the three research questions mentioned at the beginning of this Section. Notwithstanding, many questions concerning imagination remain — and many new questions pop up as a result of my analysis. I next discuss these open questions and suggest directions for future research.

Directions for future research

(i) On the conceptual geography of imagination and allied concepts. As I mentioned in the methodological preliminaries (Section 2.2) in Chapter 2, in this Chapter I have dealt with, and explicated, *many* concepts that are surrounded with controversy: not only imagination, but also visualisation, supposition, conceiving, and so on. Many different, competing and often incompatible accounts are available in the literature about every single one of these concepts that I have explicated: entire monographs have been written about nearly every concept that I discussed and explicated in this Chapter.

My proposed explications are not in agreement with *all* these accounts. This would have been impossible to achieve. But it was not my aim to propose explications that are in agreement with all accounts available in the literature. My aim, rather, was to create a *coherent* and *consistent* network of concepts which are undeniably closely related but which have never been explicitly related to each other to the extent that I did in this Chapter. In doing so, I aimed to strike a balance in my explications between (i) the six Carnapian requirements for explications mentioned in Section 2.2, p.20, and (ii) which ideas seem supported by *most* authors on the concept under investigation.

The only other conceptual geography of imagination and allied con-

cepts that I am aware of was introduced by Salis and Frigg (2020). Meynell (2021, p. 2) however argued that Salis and Frigg's conceptual network was "both unmotivated and unconvincing". I agreed with Meynell and attempted to provide an alternative that was better motivated and more convincing than Salis and Frigg's. But I will not have convinced everyone — or perhaps not *any*one (except myself). Here thus lies a fruitful direction for future research: to improve on the conceptual geography of imagination and allied concepts that I have provided in this literature. Perhaps one may find a *different* way of explicating and (coherently) connecting all the concepts that I have explicated and connected in this Chapter. Having rivaling conceptual geographies of imagination available in the literature will surely help us better understand imagination.

(ii) On the notion of voluntariness. It is widely agreed that imagination is *voluntary* to a significant extent. As I emphasised throughout this Thesis, however, the notion of "voluntariness" must be handled with extreme care. I next indicate ways in which the notion of "voluntariness" is still not fully understood, and where future research will certainly be fruitful with respect to increasing our understanding of imagination and its relation to allied concepts.

Imagination is characterised by a *voluntary attitude*. (I argued that this attitude is the voluntary attitude of *acceptance*.) Types of *proposition*-imagination that do not necessarily involve sensory content — i.e. conceiving, supposition, counterfactual thought — even seem *entirely* voluntary: both in content *and* in attitude. Nonetheless, even here our imagination — more specifically, the inferences that we make *in* our imagination — is bounded by many factors, such as the epistemic aim and background beliefs of the imaginer, and other subjective factors, see e.g. Canavotto et al. (2022). Alongside this, I discussed how a voluntary choice of *proposition*, and even a voluntary choice of *topic*, to imagine *under-determines* the content of our mental state of imagination. This means that *not all content* of our mental states of imagination is voluntary, not even the *propositional* content. Much work is done recently on providing *logics*

for imagination, which aim to capture in which ways the content of our proposition-imagination is and is not bounded, see e.g. Özgün and Schoonen (2022); Berto and Jago (2019); Berto (2017, 2021, 2022, 2023). This is a step in the right direction, as it provides provisional frameworks for how imaginers *ought* to reason. It would be a highly interesting direction for future research to test these logics empirically, to see how *actual imaginers* reason.

Perhaps more important for our understanding of the ways in which imagination is voluntary is our understanding of the ways in which *mental imagery*, i.e. *endogenous sensory content*, is voluntary. Recall (Figure 2.1) that I distinguished between *endogenous* and *exogenous* sensory content. Exogenous sensory content is *externally* caused by events via our sense organs, e.g. as in vision (2.1) and optical illusions (2.2). *Endogenous* sensory content, by contrast, is sensory content that is *not* externally caused as such (Langland-Hassan, 2020). I distinguished *imagined* mental imagery from *hallucinated* mental imagery on the basis that imagined mental imagery is *voluntary* endogenous sensory content, and hallucinated mental imagery is *involuntary* endogenous content.

But, in order to keep this distinction between hallucination and imagination sharp, I was forced to make some admittedly arbitrary choices along the way. Most notably, I decided to refer to all mental imagery that *accompanies* (Section 2.8.1) a voluntary mental state of imagination — both voluntary and involuntary accompanying mental imagery — as *imagined* mental imagery, i.e. as voluntary sensory content. But this decision was arbitrary, and it is clear that empirical research is required to find out exactly in which senses, and in which ways, mental imagery can be voluntary or involuntary; see e.g. Grealy and Lee (2011); Vyshedskiy (2020); Park et al. (2022); c.f. (Richardson, 2013). Moreover, from a conceptual point of view, I submit that the concept of *voluntariness*, with respect to mental imagery, is arguably not sharp but *vague*. Consequently, then, the distinction between imagination and hallucinations (and dreams, etc.) is also vague, not sharp. A better understanding, with a strong empirical

basis, of the ways in which mental imagery is (in)voluntary is essential for understanding imagination and its relation to allied concepts.

5.1.2 Chapter 3: Knowledge Through Imagination

In this second main Chapter, I dealt with the following research question, which I called the Question of Knowledge Through Imagination:

➢ Is imagination a source of knowledge of the natural world?

I began by distinguishing four ways in which imagination is often discussed as a (potential) source of knowledge (Section 3.2): (i) imagination as a source of *quasi-perceptual* knowledge, (ii) imagination as a source of *practical* knowledge, (iii) imagination as a source of *modal* knowledge, and (iv) imagination as *essential for* other sources of knowledge. I noted that these four ways in which imagination can function as a source of knowledge are inter-related, but that important differences pertain to the types of imagination involved and the type of knowledge gained in each. I then made clear that my analysis focuses on (i) imagination as a source of quasi-perceptual knowledge, which is arguably the most controversial way in which imagination functions as a source of knowledge.

I next explicated the concept of *quasi-perception* (3.7), which denotes both *acts of imagination* (3.1) and *episodic memories* (3.5). I discussed at length the similarities and differences between quasi-perception and 'ordinary' perception (Section 3.3.1). Following Dorsch (2016b), in analogy to a two-step reconstruction of how 'ordinary' perceptual beliefs are formed (3.8), I then put forward a *two-step schema* for the rational determination

of quasi-perceptual beliefs (3.9) (Section 3.3.2):

> **The two-step schema for quasi-perceptual beliefs:**
> (1) On the basis of quasi-perceiving (concrete observable) entity ε, proposition q with topic ε comes to mind;
> (2) On the basis of the *meta-belief* that the quasi-perceived scenario accurately represents the natural world, the propositional attitude of *belief* is adopted to q. (3.9)

To motivate this two-step schema, I discussed three examples where quasi-perceptual beliefs are obtained (Section 3.3.3), and I argued why the formation of a quasi-perceptual belief necessarily requires *meta-beliefs* about the accuracy of our quasi-perceptions, which are involved in step 2 in the two-step schema for quasi-perceptual beliefs (Section 3.3.4).

I then discussed how quasi-perceptual beliefs are justified. I first provided an explicit criterion for the justification of quasi-perceptual beliefs (3.10) (Section 3.4.1). I then discussed the Constraint Claim (3.11), which is the widely-endorsed claim that imagined manipulation of a quasi-perceived scenario is truth-preserving (condition (iii) in (3.10)) if the content of our imagination is *properly constrained in a reality-oriented way* (Section 3.4.2).

The Constraint Claim was challenged by Kinberg and Levy (2022), who argued that it gives rise to a dilemma (3.12). The dilemma ran roughly as follows. The content of an act of imagination is either (I) deliberately constrained or (II) indeliberately constrained. Horn (I): if an act of imagination is deliberately constrained, then it may yield justified quasi-perceptual beliefs, but the beliefs are not justified *in virtue of imagination*. Horn (II): if an act of imagination is indeliberately constrained, then it never yields justified quasi-perceptual beliefs. I argued first that the literature on scientific thought experiments has taught us that Horn (I) is false (Section 3.4.4). I next argued that Horn (II) is also false, because an indeliberately-constrained act of imagination can yield justified

beliefs if the imaginer is an *expert* in the imagined topic; I formulated this claim as the Reliability Claim (3.13) (Section 3.4.5). Lastly, I argued that the dilemma put forward by Kinberg and Levy (2022) is a false dilemma, as our acts of imagination are typically constrained by non-trivial *combinations* of deliberate and indeliberate constraints, which may interact in epistemologically interesting — and epistemically valuable — ways (Section 3.4.6). I next reviewed a similar dilemma (3.14) discussed — and argued against — by Miyazono and Tooming (2023a). I concluded that imagination can indeed contribute to the justification of quasi-perceptual beliefs (Section 3.4.7).

Finally, I discussed what it means to say that "imagination is a source of quasi-perceptual knowledge" (Section 3.5). I distinguished three ways in which imagination may function as a source of quasi-perceptual knowledge: (i) as a *basic* source of knowledge (3.15), (ii) as a *crucial* source of knowledge (3.16), and (iii) as source of *otherwise-inaccessible* source of knowledge (3.17). I concluded (i) that imagination is *not* a basic source of knowledge (Section 3.5.1), (ii) that imagination *is* a crucial source of knowledge in at least three different ways (Section 3.5.2), and (iii) that imagination is even a source of otherwise-inaccessible knowledge (Section 3.5.3).

Directions for future research

(i) On imagination as a source of otherwise-inaccessible knowledge. I concluded this Chapter by describing how imagination is a source of otherwise-inaccessible knowledge (Section 3.5.3). I provided one (type of) example, but I also noted that I am currently unaware of examples of other types where imagination is a source of otherwise-inaccessible knowledge. It would be very helpful for understanding in which way imagination is a source of knowledge, to have clear-cut examples of other types than the one I provided where imagination is a source of otherwise-inaccessible knowledge. As I indicated in Section 3.5.3, having such examples will increase our understanding of imagination and *memory* in the process.

This, therefore, is a promising direction for future research.

(ii) On other ways in which imagination is a source of knowledge. In Section 3.3 of Chapter 3, I distinguished four ways in which imagination arguably functions as a source of knowledge: (a) imagination as a source of *quasi-perceptual* knowledge, (b) imagination as a source of *practical* knowledge, (c) imagination as a source of *modal* knowledge, and (d) imagination as *essential for* other sources of knowledge. The focus of my analysis was on (a) imagination as a source of quasi-perceptual knowledge. But these four ways (a)–(d) are inter-related, and it is surely a promising direction for future research to bring the results of my analysis pertaining to (a) into more direct connection in the other three ways (b)–(d) in which imagination arguably functions as a source of knowledge. This also relates to the previous suggested direction for future research: intuitively, imagination may be a source of otherwise-inaccessible *practical knowledge* (e.g. increasing one's athletic ability through mental simulation for activities that one is unable to prepare for via other means) or a source of otherwise-inaccessible *modal knowledge* (e.g. knowledge of impossibilities such as the impossibility of a mountain without a valley, as Hume famously argued).

(iii) On other epistemic products of imagination. I have discussed imagination as a source of (quasi-perceptual) *knowledge*, which is the most extensively researched and, arguably, the most controversial epistemic product of imagination. But imagination functions as a source of more epistemic products than just knowledge. Notably, imagination is a source of *increased understanding* (Stuart, 2015, 2017), and, relatedly, imagination can *instigate conceptual change* (Kuhn, 1977; Steier and Kersting, 2019; Kersting et al., 2021). Moreover, processes of increasing understanding and instigating conceptual change are arguably 'less linear' than obtaining knowledge, in the sense that, it has been argued by Lombrozo (2020), even 'incorrect uses of imagination' may increase understanding in the long term. Further research on how imagination is a source of such epistemic products other than knowledge is undoubtedly a

promising direction of research. Additionally, understanding how imagination e.g. yields increased understanding or instigates conceptual change will be helpful in the context a *hot topic* in philosophy of science where the concept of imagination is highly relevant: the aesthetics of science; e.g. Brady (1998); Ivanova and French (2020).

5.1.3 Chapter 4: Scientific Thought Experiments

In the third and final main Chapter of this Thesis, *Scientific Thought Experiments*, I dealt with the following two research questions:

- What are scientific thought experiments (STEs)?
- What, and how, can we learn by performing STEs?

To provide a full-fledged answer to these two research questions, I proposed a novel full-fledged philosophical account of STEs, which is explicitly built on the recently-developed fiction view of models, and which improves on recent proposals by Meynell (2014) and Sartori (2023).

I began by discussing the two research questions mentioned above, which I called the "two core questions concerning STEs": (I) what are STEs, and (II) what, and how, do we learn by performing STEs (Section 4.2.1). I then introduced two example STEs at length (Galilei's falling bodies and Clement's Sisyphus) and briefly introduced four more STEs (Newton's bucket, Maxwell's demon, Einstein's photon-box, and Norton's dome) each of which exhibited different, characteristic features of STEs (Section 4.2.2).

With these examples in hand, I then returned to core question (I) and defined what STEs are (4.3) and explicated what it means to perform an STE (4.4) (Section 4.2.3):

> [Performing an STE] Subject S *performs* STE $= \langle \mathcal{D}, \mathcal{C}, \mathcal{E} \rangle$ iff upon engaging with the STE description $\mathcal{D}$, S reasons about imaginary scenario $\mathcal{C}$ (the topic of $\mathcal{D}$), with epistemic aim $\mathcal{E}$. (4.4)

I emphasised that both the details of the imaginary scenario of an STE and the forms of "reasoning about an imaginary scenario" mentioned in (4.4), which should be performed in order to reach the epistemic aim of the STE, are typically *under-determined* by the STE's description. This raised the question which forms of reasoning are valid ways of reaching the epistemic aim of any given STE. It seems that the epistemic aim of the STE should be *reachable* under significant variation in the under-determined features of the imaginary scenario of an STE *and* of under significant variation in the under-determined features of the reasoning process prescribed in the description of an STE. I noted that an account of STEs should be able to explain how we seem to achieve the epistemic aims of STEs so efficiently *despite* — or *because of* — the variant and invariant features of an STE.

In the subsequent Section (4.2.4), I returned to core question (II) and discussed in particular: (i) that STEs are considered to have distinct *heuristic value* and *demonstrative force*, both of which must be explained by an account of STEs, (ii) the 'paradox of thought experiments', which is much less paradoxical than it long seemed to be, and (iii) the difference between STE-beliefs and quasi-perceptual beliefs, which were the topic of the previous Chapter. I then evaluated two long-standing accounts of STEs — the argument view (Section 4.2.5) and the mental-modeling view (Section 4.2.5) — noting their respective strengths and weaknesses.

I then discussed the relation between STEs and the concept of *fiction*. I first motivated the idea that the concept of fiction is useful for understanding STEs (Section 4.3.1). I then introduced and discussed at length the theory of fiction from Walton (1990), and I provided explications for the core concepts in this theory of fiction (4.7) and its underlying conceptual schema of 'games of make-believe' (4.6).

With these explications in hand, I then discussed the account of STEs proposed by Meynell (2014), which is explicitly built on Walton's theory of fiction: Meynell reconstructs the imaginary scenario of STEs as *fictional worlds á la* Walton (Section 4.3.3). Meynell's account of STEs was a step in the right direction, but it was (by Meynell's own admission) only a

partial account of STEs. What was missing, was a consistent method for evaluating how the fictional world of an STE *relates* to the natural world, and, consequently, how the fictional world of the STE provides evidence for, or *justifies*, insight gained about the natural world by performing the STE. I then discussed the account of STEs proposed by Sartori (2023), who promised to improve on Meynell's account by introducing an explicit account of scientific representation: DEKI-representation. I concluded that Sartori succeeded only partially in improving on Meynell's account (Section 4.3.4), notably because Sartori's account *still* does not provide a consistent method for reconstructing the fictional world of an STE and for evaluating the relation between this fictional world and the natural world.

My suggestion: to complete Meynell's and Sartori's account of STEs, we should reconstruct the fictional world of an STE as a *scientific model*. To motivate this suggestion, I first discussed the close relation between STEs and scientific models, which has been noted many times before, even by Suppes (1960), but which nonetheless has been relatively under-explored in the literature on STEs (Section 4.4.1). I then introduced a recently-developed account of scientific models that is particularly well-suited for formulating a full-fledged Waltonian account of STEs: the fiction view of models (4.15) (Section 4.4.2).

Finally, with all these pieces of the conceptual puzzle in hand — Walton's theory of fiction (4.7), Walton's notion of games of make-believe (4.6) and the fiction view of models (4.15) — I proposed my own account of STEs: the fiction view of STEs (4.18). According to the proposed account, to perform a scientific thought experiment is to reason with and about scientific models (4.19). More exactly, what it means to *perform* an STE, according to the proposed fiction view of STEs, is the following:

> Subject S *performs* $\mathsf{STE} = \langle \mathcal{D}, \mathcal{F}, \mathcal{E} \rangle$ iff upon engaging with the STE description $\mathcal{D}$, S plays a game of make-believe (4.6) with $\mathcal{D}$ and reasons about the Waltonian fictional world $\mathcal{F}$ (4.7) of the STE, with epistemic aim $\mathcal{E}$. (4.19)

I formulated the proposed view as clearly as possible and noted immediate consequences of the proposed view (Section 4.5). I illustrated how STEs are analysed in light of the proposed fiction view of STEs by analysing at length Galilei's falling bodies (Section 4.5.2) and Clement's Sisyphus (Section 4.5.3). I also commented on noteworthy aspects of the other four example STEs (Section 4.5.4), thereby demonstrating the fruitfulness of the proposed account. Finally, I discussed how the proposed fiction view of STEs also provides answers to niche questions about STEs that have been posed occasionally in the literature but which have been under-illuminated, if not totally ignored, by other accounts of STEs.

Directions for future research

(i) On the scope of the proposed account. By formulating an account of thought-experiments that is explicitly built on an account of *scientific models*, I proposed an account thought experiments whose scope is limited to *scientific* thought experiments (STEs). It is an interesting direction for future research to see what my proposed account of STEs can explain about *non*-scientific thought experiments, or at least how it suggests directions for analysing them.

According to my proposed account of STEs, scientific models 'underlie' the fictional worlds of STEs. What holds together the fictional worlds of non-scientific STEs? The answer to this question will presumably differ from discipline to discipline, and perhaps it even differs on a case-by-case basis. Notwithstanding, it will be interesting to investigate what kind of philosophical analogue of 'scientific models' may perform the function that scientific models perform in my account of STEs. We may not even need an *analogue*: the concept of "model" is used in philosophy as well. It has been argued recently, for example, that the methodology of formal ethics *is* some form of model-building (Roussos, 2022; Wagner, 2023). It is thus a useful direction for future research to investigate to what extent my proposed account of STEs can apply straightforwardly — or unstraightforwardly – to thought experiments in non-scientific disciplines

such as formal ethics as well.

(ii) On the interaction between HPS and Science Education Research. In an opinion-piece on the interaction between HPS (History and Philosophy of Science) and SER (Science Education Research) with respect to the concept of imagination, myself and several co-authors argued that the interaction between HPS and SER has been too asymmetric (Alstein et al., 2022). Typically, the influence runs *from* HPS *to* SER, as newfound insight from HPS is used to improve our scientific-pedagogical methods and tools, but not the other way around, as insights from SER rarely influence HPS.

We (Alstein et al., 2022) suggested that the recently intensified research on thought experiments and other acts of imagination in SER may have an impact on our understanding of thought experiments and other acts of imagination in HPS. Much of the significant *uses* of thought-experiments in the history of science are largely beyond our epistemic reach: we can only reconstruct what happened on the basis of (often fragmented and 'Whiggish') testimony; c.f. Kind (2001) versus Kinberg and Levy (2022) on the role of imagination in Tesla's work.

Beyond this, it is rarely studied empirically how scientists *actually use* thought experiment and other acts of imagination while doing science — save a rare recent exception from (Stuart, 2023). The use of thought experiments *in the classroom* is studied empirically in SER more extensively than the use of thought experiments by scientists is studied empirically in HPS; again see Steier and Kersting (2019); Kersting et al. (2021) and the references therein. Our (Alstein et al., 2022) suggestion, then, was that we should investigate to what extent the students' use of imagination and other acts of imagination when learning new concepts *is analogous* to scientists' use of imagination when doing science and e.g. making new discoveries. If the analogy is informative to any extent, then insights from Science Education Research pertaining to thought experiments (on which there has been *a lot* of research) and other acts of imagination may begin to carry over to history and philosophy of science.

5.2 Coda

I have learned that progress in analytic philosophy often amounts to taking a proverbial step *back*, rather than forward, and learning to see the topic of your investigation from a new — hopefully *more clear* — perspective. This process indicates, hopefully, where one can increase or has increased one's understanding of the topic of their investigation, and it also shows the *limits* of their investigation. This aptly describes my experience of writing this Thesis. Through my research, I have increased my understanding of imagination, but perhaps even more so I have come to understand what I do *not* understand about imagination. Such is the way of analytic philosophy. As T.S. Eliot wrote in *Little Gidding*:

> We shall not cease from exploration
> And the end of all our exploring
> Will be to arrive where we started
> And know the place for the first time.

One must imagine Sisyphus happy.

Bibliography

Aldea, A. S. (2019). Imagination and its Critical Dimension — Lived Possibilities and an Other Kind of Otherwise. *New Yearbook for Phenomenology and Phenomenological Philosophy*, 2(XVII):Ch.14.

Alstein, P., Rijken, S., Kersting, M., and Verburgt, L. (2022). From conceptual change to scientific imagination: An interdisciplinary workshop at the crossroads of HPS and science education research. *HPS&ST Newsletter*, February 2022:24–33. http://www.hpsst.com/uploads/6/2/9/3/62931075/feboped2022.pdf. Accessed on 08-06-2022.

Arcangeli, M. (2010). Imagination in Thought Experimentation: Sketching a Cognitive Approach to Thought Experiments. In Pizzi, C., Carnielli, W., and Magnani, L., editors, *Model-Based Reasoning in Science and Technology*, pages 571–587. Springer.

Arcangeli, M. (2017). Thought Experiments in Model-Based Reasoning. In Magnani, L. and Bertolotti, T., editors, *Springer Handbook of Model-Based Science*, pages 463–493. Springer.

Arcangeli, M. (2018). The hidden links between real, thought and numerical experiments. *Croatian Journal of Philosophy*, 18(52):3–22.

Arcangeli, M. (2019). *Supposition and the Imaginative Realm: a Philosophical Inquiry*. Routledge.

Arcangeli, M. (2021). Narratives and thought experiments: Restoring the role of imagination. In Badura, C. and Kind, A., editors, *Epistemic Uses of Imagination*, pages 183–201. Routledge.

Arcangeli, M. (2023). Aphantasia demystified. *Synthese*, 201(2):31.

Aristotle (2022). *De Anima*. http://classics.mit.edu/Aristotle/soul.3.iii.html. Accessed on 29-03-2022.

Aronowitz, S. and Lombrozo, T. (2020). Learning through simulation. *Philosophers' Imprint*, 20:1–18.

Arthur, R. (1999). On thought experiments as a priori science. *International Studies in the Philosophy of Science*, 13(3):215–229.

Atkinson, D. and Peijnenburg, J. (2004). Galileo and prior philosophy. *Studies in History and Philosophy of Science Part A*, 35A(1):115–136.

Audi, R. (1994). Dispositional beliefs and dispositions to believe. *Noûs*, 28(4):419–434.

Audi, R. (1995). Memorial justification. *Philosophical Topics*, 23(1):31–45.

Audi, R. (2005). The Sources of Knowledge. In *The Oxford Handbook of Epistemology*, pages 71–94. Oxford University Press.

Bäck, A. (2005). Imagination in Avicenna and Kant. *Tópicos (México)*, 29:101–130.

Badura, C. and Kind, A., editors (2021). *Epistemic Uses of Imagination*. Routledge.

Bainbridge, W. A., Pounder, Z., Eardley, A. F., and Baker, C. I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex*, 135:159–172.

Balcerak Jackson, M. (2016). On the Epistemic Value of Imagining, Supposing, and Conceivings. In Kind, A. and Kung, P., editors, *Knowledge Through Imagination*, pages 41–60. Oxford University Press.

Barbour, J. B. and Pfister, H. (1995). *Mach's principle: from Newton's bucket to quantum gravity*, volume 6. Springer.

Bell, J. S. (2001). La Nouvelle Cuisine. In Bell, M., Gottfried, K., and Veltman, M., editors, *John S. Bell on the Foundations of Quantum Mechanics*, pages 216–234. World Scientific Publishing Co.

Beller, M. (1999). *Quantum dialogue: The making of a revolution*. University of Chicago Press.

Bennett, M. R. and Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Blackwell.

Bernecker, S. (2009). *Memory: A philosophical study*. Oxford University Press.

Bernecker, S. (2017). A causal theory of mnemonic confabulation. *Frontiers in Psychology*, 8:1207.

Berto, F. (2017). Impossible worlds and the logic of imagination. *Erkenntnis*, 82:1277–1297.

Berto, F. (2021). Turning the runabout imagination ticket. *Synthese*, 198(8):2029–2043.

Berto, F. (2022). *Topics of Thought: The Logic of Knowledge, Belief, Imagination*. Oxford University Press.

Berto, F. (2023). 'logic will get you from a to b, imagination will take you anywhere'. *Noûs (Forthcoming)*.

Berto, F. and Jago, M. (2019). *Impossible worlds*. Oxford University Press.

Bissell, C. (2007). Historical perspectives-the moniac a hydromechanical analog computer of the 1950s. *IEEE Control Systems Magazine*, 27(1):69–74.

Black, D. L. (1993). Estimation (wahm) in avicenna: The logical and psychological dimensions. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie*, 32(2):219–258.

Bokulich, A. (2001). Rethinking thought experiments. *Perspectives on Science*, 9(3):285–307.

Boniolo, G. (1997). On a unified theory of models and thought experiments in natural sciences. *International Studies in the Philosophy of Science*, 11(2):121–142.

Boudry, M. and Vlerick, M. (2014). Natural selection does care about truth. *International Studies in the Philosophy of Science*, 28(1):65–77.

Bouquiaux, L. (2008). Leibniz against the unreasonable newtonian physics. In *Leibniz: What Kind of Rationalist?*, pages 99–110. Springer.

Brady, E. (1998). Imagination and the aesthetic appreciation of nature. *The Journal of Aesthetics and Art Criticism*, 56(2):139–147.

Brendel, E. (2004). Intuition-Pumps and the Proper Use of Thought Experiments. *Dialectica*, 58:71–83.

Brendel, E. (2018). The Argument View: Are Thought Experiments Mere Picturesque Arguments? In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge Companion to Thought Experiments*, pages 281–293. Routledge.

Brewer, B. (1999). *Perception and Reason*. Oxford University Press.

Brown, D. H. (2018). Infusing Perception with Imagination. In Macpherson, F. and Dorsch, F., editors, *Perceptual Imagination and Perceptual Memory*, pages 133–160. Oxford University Press.

Brown, J. R. (1986). Thought experiments since the scientific revolution. *International studies in the Philosophy of Science*, 1(1):1–15.

Brown, J. R. (1991). *Laboratory of the Mind: Thought Experiments in Natural Sciences*. Routledge.

Brown, J. R. (2004). Why Thought Experiments Transcend Empiricism. In Hitchcock, C., editor, *Contemporary Debates in the Philosophy of Science*, pages 23–43. Blackwell.

Brown, J. R. (2007). Counter thought experiments. *Royal Institute of Philosophy Supplements*, 61:155–177.

Brown, J. R. (2013). Thought experiments. In *The Routledge Companion to Philosophy of Science*, pages 324–335. Routledge.

Brown, J. R. (2014). Explaining, seeing, and understanding in thought experiments. *Perspectives on Science*, 22(3):357–376.

Brown, J. R. and Fehige, Y. (2019). Thought Experiments. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2019 edition.

Bub, J. (2001). Maxwell's demon and the thermodynamics of computation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 32(4):569–579.

Bundy, M. W. (1922). Plato's view of the imagination. *Studies in Philology*, 19(4):362–403.

Byrne, R. M. (2016). Counterfactual thought. *Annual review of psychology*, 67:135–157.

Byrne, R. M. (2017). Counterfactual reasoning and imagination. In *International Handbook of Thinking and Reasoning*, pages 71–87. Routledge.

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. MIT Press.

Camilleri, K. (2014a). Toward a constructivist epistemology of thought experiments. *Synthese*, 191(8):1697–1716.

Camilleri, K. (2014b). Toward a constructivist epistemology of thought experiments. *Synthese*, 191(8):1697–1716.

Camilleri, K. (2015). Knowing what would happen: The epistemic strategies in galileo's thought experiments. *Studies in History and Philosophy of Science Part A*, 54:102–112.

Camp, E. (2009). Two varieties of literary imagination: Metaphor, fiction, and thought experiments. *Midwest studies in philosophy*, 33:107–130.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological bulletin*, 54(4):297.

Canavotto, I., Berto, F., and Giordani, A. (2022). Voluntary imagination: A fine-grained analysis. *The Review of Symbolic Logic*, 15(2):362–387.

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.

Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science*. Cambridge University Press.

Cartwright, N. and Le Poidevin, R. (1991). Fables and Models. *Aristotelian Society Supplementary Volume*, 65(1):55–82.

Casasanto, D. and Dijkstra, K. (2010). Motor action and emotional memory. *Cognition*, 115(1):179–185.

Chalmers, D. (2002). Does conceivability entail possibility? In *Conceivability and possibility*, pages 145–200. Oxford University Press.

Chalmers, D. J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Chignell, A. (2018). The Ethics of Belief. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2018 edition.

Chodorow, J. (1991). *Dance therapy and depth psychology: The moving imagination*. Psychology Press.

Chodorow, J. and Jung, C. (2015). *Jung on active imagination*. Princeton University Press.

Chudnoff, E. (2011). The nature of intuitive justification. *Philosophical Studies*, 153:313–333.

Chudnoff, E. and Didomenico, D. (2015). The epistemic unity of perception. *Pacific Philosophical Quarterly*, 96(4):535–549.

Clement, J. J. (1998). Expert novice similarities and instruction using analogies. *International Journal of Science Education*, 20(10):1271–1286.

Clement, J. J. (2009a). Analogical reasoning via imagery: The role of transformations and simulations. In Kokinov, B., Holyoak, K., and Gentner, D., editors, *New frontiers in analogy research*, pages 1–11. New Bulgarian University Press.

Clement, J. J. (2009b). The role of imagistic simulation in scientific thought experiments. *Topics in Cognitive Science*, 1:686–710.

Clement, J. J. (2018). Reasoning patterns in galileo's analysis of machines and in expert protocols: Roles for analogy, imagery, and mental simulation. *Topoi*, pages 1–13.

Coecke, B. and Kissinger, A. (2017). *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press.

Cohen, L. J. (1989). Belief and acceptance. *Mind*, 98(391):367–389.

Cohen, S. (2002). Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research*, 65(2):309–329.

Cohnitz, D. and Häggqvist, S. (2017). Thought experiments in current metaphilosophical debates. In *The Routledge Companion to Thought Experiments*, pages 406–424. Routledge.

Cooper, R. (2005). Thought Experiments. *Metaphilosophy*, 36(3):328–347.

Costello, D. (2018). What is abstraction in photography? *The British Journal of Aesthetics*, 58(4):385–400.

Cottrell, J. (2015). *David Hume: Imagination*. The Internet Encyclopedia of Philosophy.

Crews, F. (1995). *The Memory Wars: Freud's Legacy in Dispute*. The New York Review of Books.

Currie, G. and Ravenscoft, I. (2002). *Recreative minds: Imagination in philosophy and psychology*. Oxford University Press.

Dauer, F. W. (1999). Force and vivacity in the treatise and the enquiry. *Hume studies*, 25(1):83–99.

Davies, D. (2007). Thought experiments and fictional narratives. *Croatian Journal of Philosophy*, 7:29–45.

Davis, J. (2019). Active imagination in psychotherapy. *Recuperado el*, 20.

Dawes, A. J., Keogh, R., Andrillon, T., and Pearson, J. (2020). A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific reports*, 10(1):10022.

De Brigard, F. (2014a). Is memory for remembering? recollection as a form of episodic hypothetical thinking. *Synthese*, 191:155–185.

De Brigard, F. (2014b). The nature of memory traces. *Philosophy Compass*, 9(6):402–414.

De Brigard, F. (2017). Memory and imagination. *The Routledge handbook of philosophy of memory*, pages 127–140.

De Brigard, F. (2020). The explanatory indispensability of memory traces. *The Harvard Review of Philosophy*.

de la Torre, A. C., Daleo, A., and Garcia-Mata, I. (1999). The photon-box bohr-einstein debate demithologized. *arXiv preprint quant-ph/9910040*.

De Mey, T. (2003). The dual nature view of thought experiments. *Philosophica*, 72.

De Mey, T. (2006). Imagination's grip on science. *Metaphilosophy*, 37(2):222–239.

de Regt, H. W. (2014). Visualization as a tool for understanding. *Perspectives on Science*, 22(3):377–396.

de Regt, H. W. (2017). *Understanding scientific understanding*. Oxford University Press.

de Regt, H. W. (2020). Understanding, values, and the aims of science. *Philosophy of Science*, 87(5):921–932.

Dello Iacono, A., Ashcroft, K., and Zubac, D. (2017). Ain't Just Imagination! Effects of Motor Imagery Training on Strength and Power Performance of Athletes during Detraining. *Medicine and Science in Sports and Exercise*, 53(11):2324–2332.

Descartes, R. (2008). *Meditations on first philosophy: With selections from the objections and replies*. Oxford University Press.

Dijkstra, K., Kaschak, M. P., and Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition*, 102(1):139–149.

Dokic, J. and Arcangeli, M. (2015). The heterogeneity of experiential imagination. In Metzinger, T. K. and Windt, J. M., editors, *Open MIND*, chapter 11(T). MIND Group, Frankfurt am Main.

Dorsch, F. (2015). Focused daydreaming and mind-wandering. *Review of Philosophy and Psychology*, 6:791–813.

Dorsch, F. (2016a). Hume. In Kind, A., editor, *The Routledge Handbook of the Philosophy of Imagination*, pages 40–54. Routledge.

Dorsch, F. (2016b). Knowledge by Imagination — How Imaginative Experience can Ground Factual Knowledge. *Teorema: Revista Internacional de Filosofía*, 035(3):87–116.

Dupont, W., Papaxanthis, C., Madden-Lombardi, C., and Lebon, F. (2022). Explicit and implicit motor simulations are impaired in individuals with aphantasia. *BioRxiv*, pages 2022–12.

Earman, J. and Norton, J. D. (1998). Exorcist xiv: the wrath of maxwell's demon. part i. from maxwell to szilard. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 29(4):435–471.

Earman, J. and Norton, J. D. (1999). Exorcist xiv: The wrath of maxwell's demon. part ii. from szilard to landauer and beyond. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 30(1):1–40.

Egeland, J. (2021). Imagination cannot justify empirical belief. *Episteme*, 18(4):507–513.

Ehrenberg, W. (1967). Maxwell's demon. *Scientific American*, 217(5):103–111.

El Skaf, R. (2018). The function and limit of galileo's falling bodies thought experiment: Absolute weight, specific weight and the medium's resistance. *Croatian Journal of Philosophy*, 18(52):37–58.

El Skaf, R. (2021). Probing theoretical statements with thought experiments. *Synthese*, 199:6119–6141.

El Skaf, R. and Imbert, C. (2013). Unfolding in the empirical sciences: experiments, thought experiments and computer simulations. *Synthese*, 190:3451–3474.

El Skaf, R. and Palacios, P. (2022). What can we learn (and not learn) from thought experiments in black hole thermodynamics? *Synthese*, 200(6):434.

El Skaf, R. and Stuart, M. T. (2023). Scientific models and thought experiments: Same same but different. In *Handbook of Philosophy of Scientific Modeling*. Routledge.

Elgin, C. Z. (1996). *Considered Judgment*. Princeton University Press.

Elgin, C. Z. (2010). Telling instances. In *Beyond mimesis and convention: Representation in art and science*, pages 1–17. Springer.

Elgin, C. Z. (2014). Fiction as thought experiment. *Perspectives on Science*, 22(2):221–241.

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

Engel, P. (1998). Believing, holding true, and accepting. *Philosophical Explorations*, 1(2):140–151.

Epstude, K. (2018). Counterfactual thinking. *The psychology of thinking about the future*, pages 110–126.

Epstude, K. and Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and social psychology review*, 12(2):168–192.

Fernandez, J. (2008). Memory and time. *Philosophical Studies*, 141:333–356.

Feyerabend, P. (2020). *Against method: Outline of an anarchistic theory of knowledge*. Verso Books.

Fletcher, S. C. (2012). What counts as a Newtonian system? The view from Norton's dome. *European Journal for Philosophy of Science*, 2:275–297.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological review*, 94(4):427.

Friend, S. (2020). The fictional character of scientific models. In Smith, P. G. and Levy, A., editors, *The Scientific Imagination*, pages 101–126. Oxford University Press.

Frigg, R. (2010a). Fiction and scientific representation. In *Beyond mimesis and convention: Representation in art and science*, pages 97–138. Springer.

Frigg, R. (2010b). Models and fiction. *Synthese*, 172(2):251–268.

Frigg, R. and Nguyen, J. (2016). The fiction view of models reloaded. *The Monist*, 99.

Frigg, R. and Nguyen, J. (2017a). Models and Representation. In Magnani, L. and Bertolotti, T., editors, *Springer Handbook of Model-Based Science*, pages 49–102. Springer.

Frigg, R. and Nguyen, J. (2017b). Scientific Representation Is Representation-As. In Chao, H.-K. and Reiss, J., editors, *Philosophy of Science in Practice: Nancy Cartwright and the Nature of Scientific Reasoning*, pages 149–179. Springer.

Frigg, R. and Nguyen, J. (2018). The turn of the valve: representing with material models. *European Journal for the Philosophy of Science*, 8:205–224.

Frigg, R. and Nguyen, J. (2020). *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Springer.

Frigg, R. and Nguyen, J. (2021a). Scientific Representation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition.

Frigg, R. and Nguyen, J. (2021b). Seven Myths about the Fiction View of Models. In Casini, A. and Redmond, J., editors, *Models and Idealizations in Science. Artifactual and Fictional Approaches.*, pages 133–157. Springer.

Frigg, R. and Nguyen, J. (2022). *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Springer.

Frigg, R. and Salis, F. (2020). Of Rabbits and Men: Fiction and Scientific Modeling. In Armour-Garb, B. and Kroon, F., editors, *Fictionalism in Philosophy*, pages 187–206. Oxford Scholarship Online.

Frise, M. (2021). Reliabilism's memory loss. *The Philosophical Quarterly*, 71(3):565–585.

Frise, M. (2022). You don't know what happened. In Sant'Anna, A., McCarroll, C. J., and Michaelian, K., editors, *Current controversies in philosophy of memory*, pages 244–258. Taylor & Francis.

Galilei, G. (1638). *Dialogues concerning two new sciences*. H. Crew and A. De Salvio, Trans. Dover Publications, 1954.

Galton, F. (1880). Statistics of mental imagery. *Mind*, 5(19):301–318.

Gardner, G. (2001). Unreliable memories and other contingencies: problems with biographical knowledge. *Qualitative research*, 1(2):185–204.

Gauker, C. (2021). Imagination constrained, imagination constructed. *Inquiry*, pages 1–28.

Gaut, B. (2003). Imagination and Creativity. In Gaut, B. and Livingston, P., editors, *The Creation of Art: New Essays in Philosophical Aesthetics*, page 148–173. Cambridge University Press.

Gendler, T. S. (1998). Galileo and the indispensability of scientific thought experiment. *The British Journal for the Philosophy of Science*, 49(3):397–424.

Gendler, T. S. (2000a). The puzzle of imaginative resistance. *The Journal of Philosophy*, 97(2):55–81.

Gendler, T. S. (2000b). *Thought Experiments: On the Powers and Limits of Imaginary Cases*. Garland Press (now Routledge).

Gendler, T. S. (2004). Thought experiments rethought—and reperceived. *Philosophy of Science*, 71:1152–1164.

Gendler, T. S. (2007). Philosophical thought experiments, intuitions, and cognitive equilibrium. *Midwest Studies in Philosophy of Science*, 31:68–89.

Gendler, T. S. (2008). Alief and Belief. *The Journal of Philosophy*, 105(10):636–663.

Gendler, T. S. (2010). *Intuition, Imagination, and Philosophical Methodology*. Oxford: Oxford University Press.

Gendler, T. S. and Liao, S. (2016). The problem of imaginative resistance. In Gibson, J. and Carroll, N., editors, *The Routledge Companion to Philosophy of Literature*, pages 405–418. Routledge.

Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. University of Chicago Press.

Glas, E. (1999). Thought-experimentation and mathematical innovation. *Studies in History and Philosophy of Science Part A*, 30(1):1–19.

Godfrey-Smith, P. (2007). The strategy of model-based science. *Biology and Philosophy*, 5(21):725–740.

Goldman, A. (2008). Immediate justification and process reliabilism. *Epistemology: new essays*, pages 63–82.

Gooding, D. (1992). The Cognitive Turn, or, Why do Thought Experiments Work? In Hitchcock, C., editor, *Cognitive Models of Science*, pages 45–76. University of Minnesota Press.

Gooding, D. (1993). What is Experimental about Thought Experiments? In Hull, D., Forbes, M., and Okruhlik, K., editors, *PSA 1992, Volume Two*, pages 280–290. East Lansing, MI: Philosophy of Science Association.

Gooding, D. (1994). Imaginary science. *British Journal of Philosophy of Science*, 45:1029–1045.

Govier, T. (1972). Variations on force and vivacity in hume. *The Philosophical Quarterly (1950-)*, 22(86):44–52.

Grealy, M. A. and Lee, D. N. (2011). An automatic-voluntary dissociation and mental imagery disturbance following a cerebellar lesion. *Neuropsychologia*, 49(2):271–275.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences*, 27(3):377–396.

Gruszczyński, R. (2022). Parts of falling objects: Galileo's thought experiment in mereological setting. *Erkenntnis*, 87(4):1583–1604.

Güzel, F. (2022). Science Fiction Literature as Thought Experiment: An Ethical Analysis of Michael Crichton's Prey. *JAST*, 57:27–51.

Hacking, I. (1992). Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1992:302–308.

Häggqvist, S. (1998). Thought experiments in philosophy. *Philosophical Review*, 107(3).

Häggqvist, S. (2009). A model for thought experiments. *Canadian Journal of Philosophy*, 1(39):55–76.

Häggqvist, S. (2019). Thought experiments, formalization, and disagreement. *Topoi*, 38(4):801–810.

Hanna, R. (2022). Kant's Theory of Judgment. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition.

Hart, R. L. (1965). The imagination in plato. *International Philosophical Quarterly*, 5(3):436–461.

Hasan, A. and Fumerton, R. (2022). Foundationalist Theories of Epistemic Justification. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2022 edition.

Hemmo, M. and Shenker, O. R. (2012). *The road to Maxwell's demon: conceptual foundations of statistical mechanics*. Cambridge University Press.

Hilgevoord, J. (2002). Time in quantum mechanics. *American Journal of Physics*, 70(3):301–306.

Hopkins, R. (2018). Imagining the past. *Perceptual imagination and perceptual memory*, pages 46–71.

Horowitz, T. and Massey, G., editors (1991). *Thought Experiments in Science and Philosophy*. Lanham: Rowman and Littlefield.

Howard, D. (2007). Revisiting the einstein—bohr dialogue. *Iyyun: The Jerusalem Philosophical Quarterly*, pages 57–90.

Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, Maryland: Rowman and Littlefield.

Hume, D. (1896). *A Treatise of Human Nature*. Clarendon Press.

Hume, D. (1963). *An Enquiry Concerning Human Understanding*. La Salle: Open Court Press.

Hyde, M. (2021). *Can Imagination Give Rise to Knowledge?* PhD thesis, Stockholm University.

Ichikawa, J. and Jarvis, B. (2009). Thought-experiment intuitions and truth in fiction. *Philosophical Studies*, 142(2):221–246.

Ichikawa, J. J. (2016). Modals and Modal Epistemology. In Kind, A. and Kung, P., editors, *Knowledge Through Imagination*, pages 124–144. Oxford University Press.

Iranzo-Ribera, N. (2022). Counterfactual Reasoning in Science as Make-believe. *Synthese*, 200(6):473.

Ivanova, M. and French, S. (2020). *The Aesthetics of Science: Beauty, Imagination and Understanding*. Routledge.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17(2):187–202.

Johnson-Laird, P. (1983). *Mental Models*. Cognitive science series. Cambridge University Press.

Keogh, R. and Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, 105:53–60.

Keogh, R., Pearson, J., and Zeman, A. (2021). Aphantasia: The science of visual imagery extremes. In *Handbook of clinical neurology*, volume 178, pages 277–296. Elsevier.

Kersting, M., Haglund, J., and Steier, R. (2021). A Growing Body of Knowledge: On Four Different Senses of Embodiment in Science Education. *Science and Education*, 2(30):1183–1210.

Kieran, M. and McIver Lopes, D. (2003). Introduction. In McTravers, D., Kieran, M., and McIver Lopes, D., editors, *Imagination, philosophy, and the arts.* Routledge, Spring 2022 edition.

Kilteni, K., Andersson, B. J., Houborg, C., and Ehrsson, H. H. (2018). Motor imagery involves predicting the sensory consequences of the imagined movement. *Nature communications*, 9(1):1617.

Kim, H., Kneer, M., and Stuart, M. T. (2019). The Content-Dependence of Imaginative Resistance. In Cova, F. and Rébault, S., editors, *Advances in Experimental Philosophy of Aesthetics*, pages 143–165. Bloomsbury Publishing.

Kimpton-Nye, S. (2020). Necessary laws and the problem of counterlegals. *Philosophy of Science*, 87(3):518–535.

Kinberg, O. and Levy, A. (2022). The epistemic imagination revisited. *Philosophy and Phenomenological Research*, 00:1–18.

Kind, A. (2001). Putting the Image Back in Imagination. *Philosophy and Phenomenological Research*, pages 85—109.

Kind, A. (2013). The heterogeneity of the imagination. *Erkenntnis*, 78(1):141–159.

Kind, A. (2016). Imagining under Constraints. In Kind, A. and Kung, P., editors, *Knowledge Through Imagination*, pages 145–159. Oxford University Press.

Kind, A. (2018). How Imagination Gives Rise to Knowledge. In Macpherson, F. and Dorsch, F., editors, *Perceptual Imagination and Perceptual Memory*. Oxford University Press.

Kind, A. and Kung, P., editors (2016). *Knowledge Through Imagination*. Oxford University Press.

Knott, C. G. (1911). *Life and scientific work of Peter Guthrie Tait*, volume 1. Cambridge University Press.

Kornberger, M. and Mantere, S. (2020). Thought experiments and philosophy in organizational research. *Organization Theory*, 1(3):1–19.

Kosslyn, S. M. (1980). *Image and mind*. Harvard University Press.

Kosslyn, S. M. and Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive psychology*, 9(1):52–76.

Kosslyn, S. M. and Sussman, A. L. (1995). Roles of imagery in perception: Or, there is no such thing as immaculate perception. In Gazzaniga, M. S., editor, *The cognitive neursciences*, pages 1035–1042. Cambridge, MA: MIT Press.

Kripke, S. A. (2013). *Reference and Existence: The John Locke Lectures*. Oxford University Press.

Kubricht, J., Holyoak, K., and Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, 21(10):749–759.

Kuhn, T. S. (1977). A Function for Thought Experiments. In Kuhn, T. S., editor, *The Essential Tension*, pages 240–265. University of Chicago Press.

Kujundzic, N. (1998). The role of variation in thought experiments. *International Studies in the Philosophy of Science*, 12(3):239–243.

Kulvicki, J. V. (2007). Perceptual content is vertically articulate. *American Philosophical Quarterly*, 44(4):357–369.

Lacey, S. and Lawson, R. (2013). *Multisensory imagery*. Springer.

Lackey, J. (1999). Testimonial knowledge and transmission. *The Philosophical Quarterly*, 49(197):471–490.

Lackey, J. (2005). Memory as a generative epistemic source. *Philosophy and phenomenological research*, 70(3):636–658.

Laeng, B., Bloem, I. M., D'Ascenzo, S., and Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition*, 131(2):263–283.

Laeng, B. and Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological science*, 25(1):188–197.

Langland-Hassan, P. (2012). Pretense, imagination, and belief: the single attitude theory. *Philosophical Studies*, 159(2):155–179.

Langland-Hassan, P. (2016). On Choosing What To Imagine. In Kind, A. and Kung, P., editors, *Knowledge Through Imagination*, pages 61–84. Oxford University Press.

Langland-Hassan, P. (2020). *Explaining Imagination*. Oxford University Press.

Laraudogoitia, J. P. (2013). On norton's dome. *Synthese*, 190:2925–2941.

Lattery, M. J. (2001). Thought experiments in physics education: A simple and practical example. *Science & Education*, 10:485–492.

Leff, H. S. and Rex, A. F. (1990). *Maxwell's demon: entropy, information, computing*. Princeton University Press.

Lenhard, J. (2018). Thought experiments and simulation experiments: Exploring hypothetical worlds. In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge Companion to Thought Experiments*, pages 484–497. Routledge.

Leonard, N. (2023). Epistemological Problems of Testimony. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition.

Levy, A. (2012). Models, Fictions, and Realism: Two Packages. *Philosophy of Science*, 79(5):738–748.

Levy, A. (2015). Modeling without models. *Philosophical Studies*, 3(172):333–366.

Levy, A. and Godfrey-Smith, P., editors (2020). *The Scientific Imagination: Philosophical and Psychological Perspectives*. Oxford University Press.

Lewis, D. K. (1978). Truth in fiction. *American Philosophical Quarterly*, 15(1):37–46.

Lewis, D. K. (1986). *On the Plurality of Worlds*. Blackwell.

Li, C., Wang, J., and Otgaar, H. (2020). Creating nonbelieved memories for bizarre actions using an imagination inflation procedure. *Applied Cognitive Psychology*, 34(6):1277–1286.

Liao, S. and Gendler, T. (2020). Imagination. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2020 edition.

Liao, S. and Gendler, T. S. (2011). Pretense and imagination. *Wiley Interdisciplinary Reviews*, 2(1):79–94.

Loftus, E. F. (1997). Creating false memories. *Scientific American*, 277(3):70–75.

Loftus, E. F. and Pickrell, J. E. (1995). The formation of false memories. *Psychiatric annals*, 25(12):720–725.

Lombrozo, T. (2020). "Learning by Thinking" in Science and Everyday Life . In Levy, A. and Godfrey-Smith, P., editors, *The Scientific Imagination: Philosophical and Psychological Perspectives*, pages 230–249. Oxford University Press.

Lowe, E. (2019). *More Kinds of Being. A Further Study of Individuation, Identity, and the Logic of Sortal Terms*. Wiley-Blackwell.

Mach, E. (1883/1960). *The Science of Mechanics, sixth edition*. La Salle, IL, Open Court Publishers.

Mach, E. (1897). Über gedankenexperimente. *Zeitschrift für den physikalischen und chemischen Unterricht*, 10:1–5.

Macpherson, F. and Bermudez, J. (1998). Nonconceptual content and the nature of perceptual experience. *Electronic Journal of Analytic Philosophy*, 6.

Malament, D. B. (2008). Norton's slippery slope. *Philosophy of Science*, 75(5):799–816.

Malcolm, N. (1963). *Knowledge and certainty: Essays and lectures*. Prentice Hall, Englewood Cliff.

Marage, P. and Wallenborn, G. (1999). The debate between einstein and bohr, or how to interpret quantum mechanics: From classical to quantum mechanics. In *The Solvay Councils and the Birth of Modern Physics*, pages 161–174. Springer.

Markie, P. (2005). The mystery of direct perceptual justification. *Philosophical Studies*, 126(3):347–373.

Martin, C. B. and Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75(2):161–196.

Martin, M. G. (2002). The transparency of experience. *Mind & Language*, 17(4):376–425.

Mast, F. W. and Kosslyn, S. M. (2002). Eye movements during visual mental imagery. *Trends in Cognitive Sciences*, 6(7):271–272.

Matravers, D. (2014). *Fiction and Narrative*. Oxford University Press.

McAllister, J. W. (1996). The evidential significance of thought experiments in science. *Studies in History and Philosophy of Science Part A*, 27:233–250.

McAllister, J. W. (2012). Thought experiment and the exercise of imagination in science. In *Thought experiments in science, philosophy, and the arts*, pages 11–29. Routledge.

McCarroll, C. J., Michaelian, K., and Nanay, B. (2022). Explanatory contextualism about episodic memory: Towards a diagnosis of the causalist-simulationist debate. *Erkenntnis*, pages 1–29.

McGinn, C. (2004). *Mindsight: Image, dream, meaning*. Harvard University Press.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A*, 16(3):247.

Meynell, L. (2014). Imagination and insight: a new acount of the content of thought experiments. *Synthese*, 191(17):4149–4168.

Meynell, L. (2018). Images and imagination in thought experiments. In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge Companion to Thought Experiments*, pages 498–511. Routledge.

Meynell, L. (2021). Review of *The Scientific Imagination* by Arnon Levy and Peter Godfrey-Smith, Eds. *Mind*.

Michaelian, K. (2016a). Confabulating, misremembering, relearning: The simulation theory of memory and unsuccessful remembering. *Frontiers in Psychology*, 7:1857.

Michaelian, K. (2016b). *Mental time travel: Episodic memory and our knowledge of the personal past*. MIT Press.

Michaelian, K. and Sutton, J. (2017). Memory. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2017 edition.

Michalak, J., Mischnat, J., and Teismann, T. (2014). Sitting posture makes a difference—embodiment effects on depressive memory bias. *Clinical Psychology & Psychotherapy*, 21(6):519–524.

Miller, A. I. (2002). *Inconsistent reasoning toward consistent theories*. Springer.

Miščević, N. (1992). Mental models and thought experiments. *International Studies in the Philosophy of Science*, 6:215–226.

Miščević, N. (2007). Modelling intuitions and thought experiments. *Croatian Journal of Philosophy*, 7(2):181–214.

Miščević, N. (2022). *Thought Experiments*. Springer.

Miyazono, K. and Tooming, U. (2023a). Imagination as a generative source of justification. *Noûs*, pages 1–23.

Miyazono, K. and Tooming, U. (2023b). On the putative epistemic generativity of memory and imagination. In *Philosophical Perspectives on Memory and Imagination*. Taylor & Francis.

Morgan, M. S. (2002). Model experiments and models in experiments. In Magnani, L. and Nersessian, N. J., editors, *Model-Based Reasoning: Science, Technology, Values*, pages 41–58. Springer, Boston, MA.

Morgan, M. S. (2004). Imagination and imaging in model building. *Philosophy of Science*, 71(5):753–766.

Mulder, R. A. and Muller, F. (2023). Modal-logical reconstructions of thought experiments. *Erkenntnis*, pages 1–13.

Muller, F. (2005). The deep black sea: Observability and modality afloat. *British Journal for the Philosophy of Science*, 56:61–99.

Muller, F. (2009). The Insidiously Enchanted Forest. Essay Review of *Scientific Representation* by Bas. C van Fraassen. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 40.3:268–272.

Munro, D. (2021). Remembering the past and imagining the actual. *Review of Philosophy and Psychology*, 12(2):175–197.

Murphy, A. M. L. (2020). *Thought experiments and the scientific imagination*. PhD thesis, University of Leeds.

Myers, J. (2021). The epistemic status of the imagination. *Philosophical Studies*, 178(10):3251–3270.

Myrvold, W. C. (2011). Statistical mechanics and thermodynamics: A maxwellian view. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 42(4):237–243.

Nagel, T. (1980). What is it like to be a bat? In *The Language and Thought Series*, pages 159–168. Harvard University Press.

Nagel, T. (2012). Conceiving the impossible and the mind-body problem. *Revista Română de Filosofie Analitică*, 6(1):5–21.

Nanay, B. (2010). Attention and perceptual content. *Analysis*, 70(2):263–270.

Nanay, B. (2011). Ambiguous Figures, Attention, and Perceptual Content: Reply to Jagnow. *Phenomenology and the Cognitive Sciences*, 10(4):557–561.

Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies*, 172(7):1723–1736.

Nanay, B. (2016a). Hallucination as mental imagery. *Journal of Consciousness Studies*, 23(7-8):65–81.

Nanay, B. (2016b). The role of imagination in decision-making. *Mind & Language*, 31(1):127–143.

Nanay, B. (2021). Mental Imagery. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition.

Neisser, U. (1976). *Cognition and reality*. San Fransisco, CA: M.H. Freeman.

Nersessian, N. J. (1993). In the theoretician's laboratory: Thought experimenting as mental modeling. *Proceedings of the Philosophy of Science Association*, 2:291–301.

Nersessian, N. J. (1999). Model-Based Reasoning in Conceptual Change. In Magnani, L., Nersessian, N. J., and Thagard, P., editors, *Model-Based Reasoning in Scientific Discovery*, pages 5–23. Springer.

Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. In Carruthers, P., Stich, S., and Siegal, M., editors, *The Cognitive Basis of Science*, pages 133–153. Cambridge University Press.

Nersessian, N. J. (2007). Thought experiments as mental modelling: Empiricism without logic. *Croatian Journal of Philosophy*, VII:125–161.

Nersessian, N. J. (2018). Cognitive Science, Mental Models, and Thought Experiments. In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge Companion to Thought Experiments*, pages 309–326. Routledge.

Newton, I. (1999). *The Principia: mathematical principles of natural philosophy*. Univ of California Press.

Nichols, S. (2009). The propositional imagination. In *The Routledge Companion to Philosophy of Psychology*, pages 360–369. Routledge.

Nichols, S. and Stich, S. (2000). A cognitive theory of pretense. *Cognition*, 74:115–147.

Nichols, S. and Stich, S. P. (2003). *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press.

Noordhof, P. (2002). Imagining objects and imagining experiences. *Mind & Language*, 17(4):426–455.

Norton, J. D. (1993). Seeing the laws of nature. *Metascience*, 3:33–38.

Norton, J. D. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy*, 26:333–366.

Norton, J. D. (2003). Causation as folk science. *Philosophers' Imprint*, 3(4).

Norton, J. D. (2004a). On thought experiments: Is there more to the argument? *Philosophy of Science*, 71(5):1139–1151.

Norton, J. D. (2004b). Why Thought Experiments do not Transcend Empiricism. In Hitchcock, C., editor, *Contemporary Debates in the Philosophy of Science*, pages 44–66. Blackwell.

Norton, J. D. (2008). The dome: an unexpectedly simple failure of determinism. *Philosophy of Science*, 75:786–796.

Norton, J. D. (2013). All shook up: fluctuations, maxwell's demon and the thermodynamics of computation. *Entropy*, 15(10):4432–4483.

Norton, J. D. and Roberts, B. W. (2010). Galileo's refutation of the speed-distance law of fall rehabilitated. *Centaurus*, 54(2):148–164.

Owen, D. (2008). Hume and the Mechanics of Mind: Impressions, ideas, and association. In Fate Norton, D., editor, *The Cambridge Companion to Hume*, pages 70–104. Cmabridge University Press.

Özgün, A. and Schoonen, T. (2022). Logical development of pretense imagination. *Erkenntnis*, pages 1–27.

Pacherie, E. (2000). Levels of perceptual content. *Philosophical Studies*, 100(3):237–254.

Palmerino, C. R. (2012). Aggregating speeds and scaling motions: A response to norton and roberts. *Centaurus*, 54(2):165–176.

Palmerino, C. R. (2018). Discussing what would happen: the role of thought experiments in galileo's dialogues. *Philosophy of Science*, 85(5):906–918.

Palmieri, P. (2005). 'spuntar lo scoglio più duro': did galileo ever think the most beautiful thought experiment in the history of science? *Studies in History and Philosophy of Science Part A*, 36(2):223–240.

Park, H.-D., Piton, T., Kannape, O. A., Duncan, N. W., Lee, K.-Y., Lane, T. J., and Blanke, O. (2022). Breathing is coupled with voluntary initiation of mental imagery. *NeuroImage*, 264:119685.

Pathak, A., Patel, S., Karlinsky, A., Taravati, S., and Welsh, T. N. (2023). The "eye" in imagination: The role of eye movements in a reciprocal aiming task. *Behavioural Brain Research*, 441:114261.

Peacocke, C. (1985). Imagination, possibility and experience: A berkeleian view defended. In Foster, J. and Robinson, H., editors, *Experience and Theory*, pages 19–35. Oxford University Pres.

Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature reviews neuroscience*, 20(10):624–634.

Peijnenburg, J. and Atkinson, D. (2003). When are thought experiments poor ones? *Journal for General Philosophy of Science*, 34:305–322.

Peper, E., Lin, I.-M., Harvey, R., and Perez, J. (2017). How posture affects memory recall and mood. *Biofeedback*, 45(2):36–41.

Perler, D. (2015). *The faculties: A history*. Oxford Philosophical Concepts.

Philipse, H. (2003). The phenomenological movement. In Baldwin, T., editor, *The Cambridge History of Philosophy 1870–1945*, page 477–496. Cambridge University Press.

Pitt, D. (2022). Mental Representation. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2022 edition.

Portelli, J. P. (1979). The concept of imagination in aristotle and avicenna. *MA Thesis, McGill University*.

Potters, J. and Leuridan, B. (2004). Studying scientific thought experiments in their context: Albert einstein and electromagnetic induction. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 58:1–11.

Pounder, Z., Jacob, J., Evans, S., Loveday, C., Eardley, A. F., and Silvanto, J. (2022). Only minimal differences between individuals with congenital aphantasia and those with typical imagery on neuropsychological tasks that involve imagery. *Cortex*, 148:180–192.

Prakash, C., Stephens, K. D., Hoffman, D. D., Singh, M., and Fields, C. (2021). Fitness beats truth in the evolution of perception. *Acta Biotheoretica*, 69:319–341.

Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19):699–711.

Quine, W. (1962). Paradox. *Scientific American*, 206(4):84–99.

Reichenbach, H. (1938). *Experience and prediction: An analysis of the foundations and the structure of knowledge*. University of Chicago Press.

Reiner, M. (1998). Collaborative thought experiments in physics learning. *International Journal of Science Education*, 20(9):1043–1059.

Reiner, M. and Burko, L. M. (2003). On the limitations of thought experiments in physics and the consequences for physics education. *Science & Education*, 12:365–385.

Reiner, M. and Gilbert, J. (2000). Epistemological resources for thought experimentation in science learning. *International Journal of Science Education*, 22(5):489–506.

Reiss, J. (2012). Genealogical thought experiments in economics. In Frappier, M., Meynell, L., and Brown, J., editors, *Thought experiments in science, philosophy, and the arts*, pages 191–204. Routledge.

Rex, A. (2017). Maxwell's demon—a historical review. *Entropy*, 19(6):240.

Richardson, A. (2013). *Mental imagery*. Springer.

Rijken, S. M. (2020). A Fiction View of Scientific Thought Experiments. Presented at OZSW 2020 conference (online).

Rijken, S. M. (2021a). A Fiction View of Scientific Thought Experiments. Presented at the EPSA21 conference (online).

Rijken, S. M. (2021b). A Fiction View of Scientific Thought Experiments. Presented at the BSPS21 symposium *Fiction in Science and Metaphysics* (online).

Rijken, S. M. (2021c). Teaching with Thought Experiments. Presented at the Interdisciplinary Workshop on Conceptual Change (Utrecht University).

Rodionov, V., Zislin, J., and Elidan, J. (2004). Imagination of body rotation can induce eye movements. *Acta oto-laryngologica*, 124(6):684–689.

Roussos, J. (2022). Modelling in normative ethics. *Ethical Theory and Moral Practice*, 25(5):865–889.

Rucińska, Z. and Gallagher, S. (2021). Making imagination even more embodied: imagination, constraint and epistemic relevance. *Synthese*, 199(3-4):8143–8170.

Russell, B. (1921). *The Analysis of Mind*. Allen & Unwin Limited.

Ryle, G. (1949). *The Concept of Mind*. University of Chicago Press.

Ryle, G. (1971). *Collected Papers Volume 1. Critical Essays*. Routledge.

Sage, J. (2004). Truth-reliability and the evolution of human cognitive faculties. *Philosophical Studies*, 117(1-2):95–106.

Salis, F. (2016). The nature of model-world comparisons. *The Monist*, 3(99).

Salis, F. (2019). The new fiction view of models. *The British Journal for the Philosophy of Science*.

Salis, F. (2020). Learning through the scientific imagination. *Argumenta*, 6(1):65–80.

Salis, F. (2021). Bridging the gap: The artifactual view meets the fiction view of models. In Cassini, A. and Redmond, J., editors, *Models and Idealizations in Science*, pages 159–177. Springer.

Salis, F. and Frigg, R. (2020). Capturing the Scientific Imagination. In Levy, A. and Godfrey-Smith, P., editors, *The Scientific Imagination*, pages 17–50. Oxford University Press.

Salis, F., Frigg, R., and Nguyen, J. (2020). Models and denotation. In Falguera, J. L. and Martínez-Vidal, C., editors, *Abstract Objects: For and Against*. Springer.

Sant'Anna, A., Michaelian, K., and Perrin, D. (2020). Memory as mental time travel. *Review of Philosophy and Psychology*, 11(2):223–232.

Sartori, L. (2023). Putting the 'experiment' back into the 'thought experiment'. *Synthese*, 201(2):34.

Sartre, J.-P. (1948). *The Psychology of Imagination*. Philosophical Library.

Schabas, M. (2017). Thought experiments in economics. In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge companion to thought experiments*, pages 171–182. Routledge.

Schacter, D. L., Guerin, S. A., and Jacques, P. L. S. (2011). Memory distortion: An adaptive perspective. *Trends in cognitive sciences*, 15(10):467–474.

Schilpp, P. A. (1959). *Albert Einstein: Philosopher-Scientist*. MJF Books.

Schlaepfer, G. and Weber, M. (2017). Thought experiments in biology. In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge companion to thought experiments*, pages 243–256. Routledge.

Schmidt, H.-J. (2022). Einstein's photon box revisited. *International Journal of Theoretical Physics*, 61(7):197.

Schneider, S. (2016). Introduction: Thought experiments: Science fiction as a window into philosophical puzzles. *Science Fiction and Philosophy: From Time Travel to Superintelligence*, pages 1–16.

Schwartz, D. L. and Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1):116.

Schwitzgebel, E. (2011). Belief. In *The Routledge companion to epistemology*, pages 14–23. Routledge.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(3):417–424.

Searle, J. R. (1982). The chinese room revisited. *Behavioral and brain sciences*, 5(2):345–348.

Senor, T. D. (1996). The prima/ultima facie justification distinction in epistemology. *Philosophy and Phenomenological Research*, 56(3):551–566.

Senor, T. D. (2019). Epistemological Problems of Memory. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2019 edition.

Shapiro, L. and Spaulding, S. (2021). Embodied Cognition. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition.

Shepard, R. N. and Cooper, L. A. (1986). *Mental images and their transformations*. The MIT Press.

Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703.

Shields, C. (2020). Aristotle's Psychology. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2020 edition.

Shinod, N. (2021). Why computer simulation cannot be an end of thought experimentation. *Journal for General Philosophy of Science*, 52(3):431–453.

Silova, I. (2020). The power of radical thought experiments: Reading feminist science fiction in comparative education. *Comparative Education Review*, 64(1):147–149.

Silva, J. (2020). How modular are medieval cognitive theories? In Dresvina, J. and Blund, V., editors, *Cognitive Sciences and Medieval Studies*. University of Wales Press.

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction*, 25(4):337–398.

Smith, J. (2006). Bodily awareness, imagination and the self. *European Journal of Philosophy*, 14(1):49–68.

Sorensen, R. A. (1992). *Thought Experiments*. Oxford University Press.

Sprenger, A., Lappe-Osthege, M., Talamo, S., Gais, S., Kimmig, H., and Helmchen, C. (2010). Eye movements during rem sleep and imagination of visual scenes. *Neuroreport*, 21(1):45–49.

Squire, L. R. (2009). Memory and brain systems: 1969–2009. *Journal of Neuroscience*, 29(41):12711–12716.

Squire, L. R., van der Horst, A. S., McDuff, S. G., Frascino, J. C., Hopkins, R. O., and Mauldin, K. N. (2010). Role of the hippocampus in remembering the past and imagining the future. *Proceedings of the National Academy of Sciences*, 107(44):19044–19048.

Stalnaker, R. (1984). *Inquiry*. MIT Press.

Stalnaker, R. C. (1968). A theory of conditionals. In *Ifs: Conditionals, belief, decision, chance and time*, pages 41–55. Springer.

Steier, R. and Kersting, M. (2019). Metaimagining and Embodied Conceptions of Spacetime. *Cognition and Instruction*, 2(37):145–168.

Strawson, P. (1970). Imagination and perception. In Foster, L. and Swanson, J., editors, *Experience and Theory*, page 31–54. University of Massachusetts Press.

Streminger, G. (1980). Hume's theory of imagination. *Hume Studies*, 6(2):91–118.

Stuart, M. T. (2015). *Thought Experiments in Science*. PhD thesis, University of Toronto.

Stuart, M. T. (2016). Taming theory with thought experiments: Understanding and scientific progress. *Studies in History and Philosophy of Science*, 58:24–33.

Stuart, M. T. (2017). Imagination: A sine qua non of science. *Croatian Journal of Philosophy*, 49(17):9–32.

Stuart, M. T. (2018). How Thought Experiments Increase Understanding. In Stuart, M. T., Fehige, Y., and Brown, J. R., editors, *The Routledge Companion to Thought Experiments*, pages 526–544. Routledge.

Stuart, M. T. (2019). The Rise of Chemical Thought Experiments. Presented at the MetaMetaPhysical Club (2019).

Stuart, M. T. (2020). The productive anarchy of scientific imagination. *Philosophy of Science*, 87(5):968–978.

Stuart, M. T. (2021). Towards a dual process epistemology of imagination. *Synthese*, 198:1329–1350.

Stuart, M. T. (2023). Scientists are epistemic consequentialists about imagination. *Philosophy of Science*, 90(3):518–538.

Stuart, M. T., Fehige, Y., and Brown, J. R., editors (2018). *The Routledge Companion to Thought Experiments*. Routledge.

Sulutvedt, U., Mannix, T. K., and Laeng, B. (2018). Gaze and the eye pupil adjust to imagined size and distance. *Cognitive Science*, 42(8):3159–3176.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese*, 2-3(12):287–301.

Swirski, P. (2007). *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution, and Game Theory*. Routledge.

Thomasson, A. L. (2020). If Models Were Fictions, Then What Would They Be? In Levy, A. and Godfrey-Smith, P., editors, *The Scientific Imagination: Philosophical and Psychological Perspectives*, pages 51–74. Oxford University Press.

Thompson, B. (2010). The spatial content of experience. *Philosophy and Phenomenological Research*, 81(1):146–184.

Thomson, J. J. (2004). A defense of abortion. In *Ethics: Contemporary Readings*, pages 267–274. Routledge.

Thomson, W. (1874). The kinetic theory of the dissipation of energy. *Nature*, 9:325–334.

Tidman, P. (1994). Conceivability as a test for possibility. *American Philosophical Quarterly*, 31(4):297–309.

Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction, and Scientific Representation*. Palgrave Macmillan.

Traiger, S. (2008). Hume on memory and imagination. In Radcliffe, E. S., editor, *A Companion to Hume*, pages 58–71. Blackwell.

Tuna, E. H. (2020). Imaginative Resistance. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2020 edition.

Tuomela, R. (2000). Belief versus acceptance. *Philosophical Explorations*, 2:122–137.

Urmson, J. (1967). Memory and imagination. *Mind*, 76:83–91.

Vaihinger, H. (1924). *The Philosophy of As If: A System of the Theoretical, Practical and Religious Fictions of Mankind*. Routledge and Kegan Paul.

van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press.

Van Hoeck, N., Watson, P. D., and Barbey, A. K. (2015). Cognitive neuroscience of human counterfactual reasoning. *Frontiers in human neuroscience*, 9:420.

Van Leeuwen, N. (2016). Imagination and Action. In Kind, A., editor, *The Routledge Handbook of Philosophy of Imagination*, pages 286–299. Routledge.

Van Strien, M. (2014). The norton dome and the nineteenth century foundations of determinism. *Journal for General Philosophy of Science*, 45:167–185.

Vendler, Z. (1984). *The Matter of Minds*. Clarendon Press.

Vickers, J. N. (2007). *Perception, cognition, and decision training: The quiet eye in action*. Human Kinetics.

Vyshedskiy, A. (2020). Voluntary and involuntary imagination: Neurological mechanisms, developmental path, clinical implications, and evolutionary trajectory. *Evolutionary Studies in Imaginative Culture*, 4(2):1–18.

Wagner, I. A. (2023). Ethical theories as multiple models. *Journal of Medical Ethics*, 49(6):444–446.

Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.

Walton, K. L. (1991). Précis of mimesis as make-believe: On the foundations of the representational arts. *Philosophy and Phenomenological Research*, 51(2):379–382.

Wedgwood, K. B. (1977). *The development of the concept of imagination from Plato and Aristotle to its introduction into English art educational theory*. PhD thesis, University of Warwick.

Wegner, D. and Schneider, D. (2003). The white bear story. *Psychological Inquiry*, 14:326–329.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

White, A. R. (1990). *The Language of Imagination*. Blackwell.

Wigner, E. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13:1–14.

Wilbanks, J. (2012). *Hume's theory of imagination*. Springer.

Wilde, D. (2011). *Extending body and imagination: moving to move*. Walter de Gruyter.

Williams, D. (2021). Imaginative constraints and generative models. *Australasian Journal of Philosophy*, 99(1):68–82.

Williamson, T. (2005). Armchair Philosophy, Metaphysical Modality, and Counterfactual Thinking. *Proceedings of the Aristotelian Society*, 105:1–23.

Williamson, T. (2016). Knowing by Imagining. In Kind, A. and Kung, P., editors, *Knowledge Through Imagination*, pages 113–123. Oxford University Press.

Wilson, J. (2023). Determinables and Determinates. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition.

Wiltsche, H. A. (2021). The forever war: understanding, science fiction, and thought experiments. *Synthese*, 198(4):3675–3698.

Winterbourne, A. (1985). Newton's arguments for absolute space. *Archiv für Geschichte der Philosophie*, 67.

Wittgenstein, L. (1980). *Remarks on the Foundations of Psychology. Volume 2*. University of Chicago Press.

Wyer Jr., R. S. (2007). Principles of mental representation. *Social psychology: Handbook of basic principles*, 2:285–307.

Xhignesse, M.-A. (2021). Imagining fictional contradictions. *Synthese*, 199:3169–3188.

Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, 53(1):1–42.

Yablo, S. (2014). *Aboutness*. Princeton University Press.

Yaldir, H. (2009). Ibn Sînâ (Avicenna) and René Descartes on the faculty of imagination. *British Journal for the History of Philosophy*, 17(2):247–278.

Yeates, L. B. (2004). *Thought experimentation: A cognitive Approach*. PhD thesis, University of New South Wales.

Zemach, E. (1968). A definition of memory. *Mind*, 77(308):526–536.

Zeman, A., Dewar, M., and Della Sala, S. (2015). Lives without imagery–congenital aphantasia. *Cortex*, 73:378–380.

Zeman, A. Z., Dewar, M. T., and Della Sala, S. (2016). Reflections on aphantasia. *Cortex*, 74:336–337.

Summary

This Thesis is a conceptual and epistemological analysis of *imagination* and *scientific thought experiments*. This Thesis consists of an Introduction, three main Chapters and a general Conclusion that summarises the obtained results and identifies directions for future research.

In the first main Chapter of this Thesis, *Explicating Imagination*, I propose Carnapian explications for the concept of *imagination* and many of its closely-related concepts. I begin by distinguishing imagination from *perception*, *optical illusions*, and *hallucination*. I then distinguish two *types* of imagination: *proposition*-imagination and *action*-imagination. I first propose an explication for proposition-imagination, and I discuss how this explication holds in light of — and sheds a new light on — eight 'core characteristics' that are often associated with imagination in the literature. Using this explication, I then explicate the concepts of *supposition*, *counterfactual thought*, *conceiving*, *visualisation* and *picturing* as types of proposition-imagination. I then turn to explicating the second type of imagination: action-imagination. Using this explication of action-imagination, I revisit what it means to visualise and picture actions, and I relate imagination to *memory*. Finally, I comment on characteristics aspects of imagination in practice, and I provide some brief but necessary notes on the cognitive science of imagination.

In the the second main Chapter, I discuss how imagination can function as a source of knowledge of the natural world. I begin by explicating, in contrast to 'ordinary' perception, the concept of *quasi-perception*, i.e.

the 'perception-like' mental state that we have when we *imagine* perceptions or vividly *remember* the past. I provide a two-step framework for how we obtain novel *beliefs* about the natural world on the basis of quasi-perceptions, which I call *quasi-perceptual beliefs*. I then discuss at length how quasi-perceptual beliefs are epistemically *justified*. I then discuss how *imagination* can be responsible for this justification. Finally, I distinguish and discuss several senses of the term "source of knowledge". I conclude that (i) imagination is not a so-called *basic* source of knowledge, (ii) imagination is certainly what I call a *crucial* source of knowledge, and (iii) imagination is even what I call a source of *otherwise-inaccessible* knowledge.

In the third and final main Chapter of this Thesis, *Scientific Thought Experiments*, I discuss what scientific thought experiments (STEs) are and what, and how, we learn by performing them. I introduce several example STEs, each of which serve to illustrate important characteristics of STEs. I then elaborate on the two research questions mentioned-above, and I discuss two long-standing accounts of STEs — the argument view and the mental-modeling view — indicating their strengths and weaknesses. I then introduce the theory of fiction from Walton (1990) and discuss two recently proposed accounts of STEs that are explicitly built on this theory of fiction. To improve on these recent proposals, I then introduce the *fiction view of models*, which I use to formulate a full-fledged account of STEs: the *fiction view of scientific thought experiments*. I argue in favor of the proposed account by analysing at length several example STEs and by indicating how the proposed account provides answers to a wide variety of question about STEs that have been posed in the literature.

Samenvatting

Dit proefschrift is een conceptuele en epistemologische analyse van de *voorstelling*[129] en van *wetenschappelijke gedachte-experimenten*. Dit proefschrift bestaat uit een introductie, drie hoofdstukken, en een conclusie waarin ik de behaalde resultaten samenvat en richtingen voor toekomstig onderzoek beschrijf.

In het eerste hoofdstuk van dit proefschrift, *Explicating Imagination*, stel ik Carnapiaanse *explicaties* (kortweg: combinaties van noodzakelijke en voldoende voorwaarden) voor voor het concept *voorstelling* (Engels: *imagination*) en nauwverwante concepten. Ik begin met het maken van een onderscheid tussen voorstelling en *perceptie*, *optische illusies* en *hallucinatie*. Vervolgens onderscheid ik twee *types* voorstelling: *propositie*-voorstelling en *actie*-voorstelling. Ik stel eerst een explicatie voor propositie-verbeelding voor, en ik bespreek ik hoe deze explicatie zich sterk houdt in het licht van — en een nieuw licht werpt op — acht 'kernkenmerken' die in de literatuur vaak met voorstelling worden geassocieerd. Met behulp van deze explicatie expliceer ik vervolgens de concepten *veronderstelling* (Engels: *supposition*), *tegenfeitelijke gedachtes* (Engels: *counterfactual thought*), het vormen van een *denkbeeld* (Engels: *conceiving*), *visualisatie* en *inbeelding* (Engels: *picturing*) als typen propositie-

[129] Het Engelse "imagination" kan in het Nederlands evenwel vertaald worden als *verbeelding* of *voorstelling*. Ik heb voor de laatste optie gekozen, mede omdat ik het woord "verbeelding" (net als het Engelse "imagination") misleidend vind: wij kunnen ons dingen *voorstellen* zonder het ook *in te beelden*. Daarnaast hangt het woord "verbeelding" vaak onterecht samen met *onwaarheid*. Wat wij ons voorstellen, kan prima waar zijn. Zie ook voetnoot 16 in Hoofdstuk 1, pagina 11.

voorstelling. Vervolgens richt ik mij op de explicatie van het tweede type voorstelling: actie-voorstelling. Met behulp van deze explicatie van actie-voorstelling bekijk ik wat het betekent om acties (*activiteiten*) te visualiseren en in te beelden, en breng ik voorstelling in verband met *geheugen*. Ten slotte geef ik commentaar op kenmerkende praktische aspecten van de voorstelling en maak ik enkele korte doch noodzakelijke opmerkingen over het cognitief-wetenschappelijk perspectief op ons voorstellingsvermogen.

In het tweede hoofdstuk, *Knowledge Through Imagination*, bespreek ik hoe ons voorstellingsvermogen kan fungeren als een bron van kennis van de natuurlijke wereld. Ik begin met het expliceren van het concept van *quasi-perceptie*, d.w.z. de 'perceptie-achtige' mentale toestand die we hebben als we percepties *voorstellen*, of als we levendig het verleden *herinneren*, welke in tegenstelling staat tot 'gewone' zintuigelijke perceptie. Ik formuleer een tweestapsraamwerk voor hoe we nieuwe *overtuigingen* (Engels: *beliefs*) over de natuurlijke wereld verkrijgen op basis van quasi-percepties, welke ik *quasi-perceptuele overtuigingen* noem. Vervolgens bespreek ik uitvoerig hoe quasi-perceptuele overtuigingen epistemisch *gerechtvaardigd* kunnen zijn, en ik bespreek hoe ons *voorstellingsvermogen* verantwoordelijk kan zijn voor deze rechtvaardiging. Ten slotte onderscheid en bespreek ik verschillende betekenissen van de term 'kennisbron'. Ik concludeer dat (i) ons voorstellingsvermogen geen zogeheten *basis*-kennisbron is, (ii) ons voorstellingsvermogen zeker een (in mijn woorden) *cruciale* kennisbron is, en (iii) ons voorstellingsvermogen zelfs een bron van (in mijn woorden) *anderzijds-ontoegankelijke* kennis is.

In het derde en laatste hoofdstuk van dit proefschrift, *Scientific Thought Experiments*, bespreek ik wat wetenschappelijke gedachte-experimenten (WGEn) zijn en wat, en hoe, we leren door ze uit te voeren. Ik introduceer verschillende voorbeelden van WGEn, die elk dienen om belangrijke kenmerken van WGEn te illustreren. Vervolgens ga ik dieper in op de twee hierboven genoemde onderzoeksvragen, en bespreek ik twee bekende filosofische visies op WGEn — de argumentatievisie en de mentale-modelleringsvisie — waarbij ik hun sterke en zwakke punten aanduid.

Vervolgens introduceer ik de fictietheorie van Walton (1990) en bespreek ik twee onlangs voorgestelde visies op WGEn, die expliciet op deze fictietheorie zijn gebaseerd. Om deze recente voorstellen te verbeteren, introduceer ik vervolgens de recent-ontwikkelde *fictie-visie van modellen*, welke ik gebruik om een volwaardige visie van WGEn te formuleren: de *fictie-visie van wetenschappelijke gedachte-experimenten.* Ik onderbouw mijn voorgestelde visie door een aantal voorbeelden van WGEn uitvoerig te analyseren en door aan te geven hoe mijn voorgestelde visie antwoorden geeft op een breed scala aan vragen over WGEn die door de jaren heen in de literatuur zijn gesteld.

About the Author

Sam Rijken (1993) holds an MSc (*cum laude*) in philosophy of science and a BSc in physics and astronomy at Utrecht University. He has presented academic work in philosophy of physics and general philosophy of science at many workshops and conferences. He taught many courses in physics, philosophy of science and philosophy proper. He enjoys public speaking and has given public talks in e.g. churches, cinemas and festivals on a variety of topics such as: science, religion, education, free will, morality, the meaning of life, friendship, imagination, thought experiments, artificial intelligence, The Matrix, YouTube and our relation to technology.

Contrary to what the form or content of this Thesis might suggest, Sam does not take life too seriously. He no longer thinks that life, by itself, is a philosophical problem. Instead of philosophizing about life, he rather enjoys sports, chess, yoga and making music. He also loves dogs.