Thesis

DPhil Philosophy

Criss-Crossing Creativity:

On Creativity, Wittgenstein, and Artificial Intelligence

Jonas Bozenhard

The Queen's College
University of Oxford

Table of Contents

Acknowledgements

First, I would like to sincerely thank my supervisors Prof Stephen Mulhall and Prof Ben Morgan for their guidance and support during my DPhil. It is an extraordinary privilege to have had the opportunity to study under their supervision. Thank you very much for all the inspiring, encouraging, insightful, and often challenging meetings and discussions, without which I would not have been able to write this thesis.

Gratitude is also due to my Transfer and Confirmation of Status examiners Prof Alice Crary, Prof Adrian Moore, Prof Barry Murnane, and Prof Ritchie Robertson for the constructive feedback and the stimulating conversations during the Transfer and Confirmation examinations and beyond.

Moreover, I would like to thank my family and friends. This thesis was conceived and written in turbulent times, with crisis following crisis. Thanks to the emotional support of my family and friends, I was able to complete this project, despite the challenging circumstances.

Finally, I am utterly grateful to the Arts and Humanities Research Council for receiving an AHRC Doctoral Training Partnership Studentship and to The Queen's College and the Waverley Fund for obtaining a Waverley Scholarship. I would also like to thank the University of Oxford for granting me a Humanities Covid-19 Scholarship Extension, as well as the Royal Institute of Philosophy for being awarded a Royal Institute of Philosophy Bursary.

To create the cover picture, I used craiyon.com to generate images which I then edited by using image processing software.

Abstract

Until lately, the view prevailed that creativity is the last stronghold of human supremacy, unconquerable by even the most intelligent machines. Yet, recent advances in artificial intelligence (AI) cast doubt on this cherished conviction: there are now programmes that generate music in the style of famous composers or produce paintings that are displayed in renowned galleries. In the face of these awe-inspiring algorithmic achievements, this thesis pursues two main goals. First, it devises a new set of conceptual tools for analysing human creativity. It presents a novel practice-based account of creativity which captures the multifacetedness and distributed character of this phenomenon. Ludwig Wittgenstein's later philosophy, especially his remarks on rule-following, serve as a crucial source of inspiration in this context. The framework that will be developed overcomes the shortcomings of existing definitions and theories that pervade the philosophical and scientific literature on creativity. Secondly, this thesis examines whether and to what extent creativity can legitimately be attributed to machines in the era of AI. It engages critically with influential objections to the notion of computational creativity and assesses the creative potential and limitations of state-of-the-art AI systems. The proposed approach challenges strong forms of anthropocentrism as well as the hyperbole evoked by recent developments in deep learning. So, the present work is not just a contribution to the emerging philosophy of creativity, but also to the blossoming philosophy of AI. The thesis concludes with a speculative exploration of possible conceptual futures for "creativity" in the face of potential future machines with superhuman creative powers.

1. Criss-Crossing a Wide Field: Creativity, Wittgenstein, and Artificial Intelligence

For a long time, humanity has taken pride not just in being the peak of creation, but also in being the peak of creativity. Creativity allowed us not only to create timeless art and unravel some of the most profound mysteries of the universe. It also allowed us to build machines that are able, or at least appear to be able, to do things we were convinced could only be done by humans. This is largely due to advances in research areas that are commonly bundled under the label of "artificial intelligence" or "AI". In order to allay any concerns that may arise from this situation and to protect ourselves against looming narcissistic injuries, we tend to reassure ourselves of what enabled us to build such machines in the first place, our creativity. There is a pervasive narrative according to which machines, even if they become more and more powerful and seemingly intelligent, cannot be creative: "Human creative achievement[…] will not succumb to advances in artificial intelligence" (Kelly 2019: 70). So, until lately, the view prevailed that creativity is the last stronghold of human supremacy, unconquerable by even the most intelligent machines.

However, recent developments in AI cast doubt on this cherished conviction: there are now programmes that "compose" music in the style of Bach or generate paintings that are displayed in renowned galleries. In the face of these awe-inspiring algorithmic achievements, we are confronted with the question whether and to what extent machines can be creative. This, in turn, raises questions about the nature and workings of what we call "creativity". While the ascription of creativity to AI systems is a contentious matter, there is little doubt that humans can be creative. So, before we attend to the issue of machine creativity, we first require a better understanding of human creativity.

Creativity is a crucial feature of human life. Nevertheless, while Western philosophers have been thinking about human capacities such as reason, knowledge, or consciousness for centuries, they have tended to neglect the phenomenon of creativity: there is no *Enquiry Concerning Human Creativity*, no *Critique of the Power of Creativity*, no *Phenomenology of*

Creativity. However, over the last ten years, creativity has received more and more attention in the context of academic philosophy – to the extent that one can speak of an emerging "philosophy of creativity" today. This thesis is, in part, a contribution to this new branch of philosophical inquiry. However, it will be shown that the most influential philosophical and scientific theories of creativity exhibit conceptual shortcomings – a situation which calls for a new way of making sense of the phenomenon at issue. Consequently, this thesis pursues two main goals: it (a) devises a new set of conceptual tools for analysing human creativity and (b) examines whether and to what extent creativity can legitimately be attributed to machines in the era of AI.

In the scientific and philosophical literature on (human) creativity, there is a strong tendency to conceive of creativity in terms of rules: spectacular examples of creativity change the rules of a certain domain by breaking with them. Yet, there is disagreement about what it means to break – and, corresponding to that, to follow – a rule. To gain a better understanding of this phenomenon, we will draw on the conceptual resources provided by Ludwig Wittgenstein's later writings, especially in *Philosophical Investigations*, which feature arguably the most influential discussion of what it means to follow a rule in modern philosophy. And even beyond his reflections on rule-following, Wittgenstein's later philosophy will serve as a crucial source of inspiration in thinking about human creativity.

Most theories and definitions in the scientific and philosophical literature fail to do justice to the multifacetedness and diversity of the phenomenon of creativity. By drawing on Wittgensteinian ideas as well as contemporary philosophy of art and science, we will provide a new vocabulary for discerning the various facets of the concept of "creativity". This will allow us to explore the term's conceptual landscape, to strike a Wittgensteinian note, "criss-cross in every direction" (PI: vii).[1]

[1] When referring to Wittgenstein's *Philosophical Investigations* (1986), I use the abbreviation "PI".

However, when one traverses the rugged and labyrinthine landscape of contemporary Wittgenstein scholarship, one might get the impression that there is not just one Wittgenstein. It is as if there were, besides the early Wittgenstein, not only a single late, but numerous late Wittgensteins: "Countless interpretations now abound" (Conant 2004: 867) – a circumstance which is due to the fragmented and, at times, enigmatic style of *Philosophical Investigations* and the text's complicated editorial history. Yet, here, the main concern is not an interpretative one, but an argumentative or systematic one. Therefore, even if the following investigations are inspired by Wittgenstein, there is no claim that they fully reflect Wittgenstein's thoughts. Rather, we appeal to Wittgenstein's remarks on rule-following and key concepts of his later philosophy in a way that seems, from our perspective, most accurate on argumentative grounds with respect to the systematic questions in focus – irrespective of whether we interpret Wittgenstein correctly in doing so. Would the later Wittgenstein approve of this way of referring to his philosophical work? In light of the following statement from the Preface to *Philosophical Investigations*, he probably would: "I should not like my writing to spare other people the trouble of thinking. But, if possible, to stimulate someone to thoughts of his own" (PI: viii).

Wittgenstein's later philosophy will also play a key role when grappling with the issue of machine creativity. Several passages of this later work revolve around the question of whether "it [is] possible for a machine to think" (Wittgenstein 1964: 47). So, the engineer-turned-philosopher Wittgenstein can be read as a philosopher of AI *avant la lettre*. In fact, Wittgenstein played an insufficiently acknowledged role in the formation of AI as a systematic research area. For none other than Alan Turing, who is widely considered the father of the computer and AI research, studied with Wittgenstein at Cambridge in 1939. Despite some considerable disagreements over issues in the philosophy of mathematics (as documented in Wittgenstein (1978)), several aspects of Turing's work exhibit his teacher's influence (Neumaier 1989: 132f.; Floyd 2017). Apart from that, Wittgenstein's thought also provided a significant impetus for many other AI pioneers, including his student Margaret Masterman and

her ground-breaking work on machine translation.[2] At the same time, some of the most vehement critics of the enterprise we now call AI – such as Hacker (2019), Shanker (2002), and Dreyfus (1972) – regard the late Wittgenstein as an intellectual ally whose philosophy challenges or even undermines the pursuit of constructing "thinking machines" – and by extension the endeavour of building "creative machines". Due to these striking reception-specific tensions, exploring the idea of artificial creativity in dialogue with the late Wittgenstein promises to provide useful insights.

Before we start the argument, here is a brief overview of the individual chapters. *Chapter 2* sets the scene by situating the project within a wider theoretical and historical context. As highlighted above, "creativity" is, first and foremost, seen as a human capacity. Section 2.1 elaborates on the historical underpinnings of this human-centred paradigm of creativity. Section 2.2 engages critically with definitions of "creativity" that permeate current philosophical and scientific debates about creativity and exposes their limitations. Section 2.3 attends to influential theories of creativity, including Darwinian, combinatorial, computational, and socio-contextual models. It will be shown that these theoretical accounts, although they point to important aspects of creativity, draw an incomplete picture of the phenomenon at issue.

Chapter 3 provides the basis for our own account of "creativity". Section 3.1 considers how to overcome the shortcomings of the definitions of – or, more precisely, the modes of defining – "creativity" that have been criticised in the previous chapter. Section 3.2 then formulates requirements for a satisfying account of creativity. The remaining sections concentrate on the above-mentioned issue of rule-following: while section 3.3 criticises reductionist takes on the subject matter, section 3.4 presents a Wittgenstein-inspired, anti-reductionist, practice-based alternative which shapes the consequent orientation of the inquiry.

[2] For a fuller picture of Wittgenstein's impact on early AI research, see Liu (2021).

Chapter 4 presents a novel framework of creativity which captures the multifacetedness and distributed character of this phenomenon and overcomes the shortcomings of existing theories and definitions. Section 4.1 characterises the approach that will be pursued to achieve this, one which revolves around the notion of what Wittgenstein refers to as a "perspicuous representation" (PI: §122). Sections 4.2 and 4.3 introduce a framework which helps us to perspicuously represent "creativity" in its synchronic and diachronic dimensions. Section 4.4 takes a closer look at some important differences between artistic and scientific creativity by drawing on this framework.

Chapter 5 finally turns to the question of artificial creativity. Sections 5.1 and 5.2 scrutinise objections to the idea that machines can be creative that have proven highly influential in contemporary academia and beyond. Sections 5.3 and 5.4 then attend to seemingly Wittgenstein-inspired points of criticisms. On this basis, section 5.5 presents a practice-based account of artificial creativity. Section 5.6 uses this framework to assess the creative potential and limitations of cutting-edge AI systems today. The proposed approach to computational creativity challenges strong forms of anthropocentrism as well as the hyperbole evoked by recent developments in deep learning.

The coda considers the possible formation of a future posthuman paradigm of creativity that may supersede the current human-centred paradigm. This raises the question of how to conceive of creativity in AI terms, i.e., in a way that is not derivative of the *conditio humana,* but true to the *conditio intelligentiae artificialis,* as it were. The coda concludes with some speculative reflections on possible conceptual futures for "creativity" in the face of potential future machines with superhuman creative powers.

2. Creativity in History and Theory

2.1 Who Can Be Creative? Two Historical Paradigms

Today, it is uncontroversial to say that humans can be creative. In modern, secular, capitalistic societies, creativity is even widely regarded as "more or less compulsory in an increasing number of areas of life" and celebrated as "the highest achievable good" (Osborn 2003: 507f.). Some even speak of a prevailing "creativity imperative" (Reckwitz 2017: 3; 2018). So, in contemporary Western culture, it is a truism to say that humans can be, are, and need to be creative. However, a couple of centuries ago, attributing "creativity" to humans would have been considered absurd, nonsensical.

In part, this reflects the fact that the term "creativity" is of quite recent origin. It is an invention of the late nineteenth century. The English historian and literary scholar Adolfus William Ward apparently introduced the word "creativity" in his 1875 book *History of English Dramatic Literature*, when describing Shakespeare's poetry (1875: 506). The term "creativity" is a nominalisation of the English verb "to create" which derives from the Latin verb *creare* ("bringing something forth") and the cognate noun *creatio* which can, in turn, – according to the standard etymological view – be traced back to the assumed Indo-European root *ker*, *kere* ("to grow"). Even though its first recorded use dates to the second half of the nineteenth century, it would take more than half a century for "creativity" to become a household word and to be adopted by other European languages (Weiner 2000: 8, 89; Sawyer 2012: 19; Glăveanu & Kaufman 2019: 10).

However, the concept of "creativity" did not emerge in a vacuum. (In other words, its invention was no *creatio ex nihilo*.) Rather, it draws on various conceptual predecessors from different eras that still shape how we think and talk about "creativity" in the present. To put it differently, the concept of "creativity" has a long and rich "pre-history" which spans more than two millennia. However, this conceptual "pre-history" tells a vastly different story of *who* can be creative. (The pre-history of the notion of "creativity" is inextricably linked to European

intellectual history. So, despite the notion's more international appeal in today's globalised world, its pre-history is predominantly a Western one. Therefore, the story that will be told in this chapter concentrates on Western ideas and concepts.[3])

In the following, I will outline two consecutive historical paradigms for who or what is considered "creative". The first one is mainly a phenomenon of the past (dominating the largest portion of the concept's pre-history); the second one a phenomenon of the past (dating back to the concept's pre-history and shaping the largest part of its history) and the present.

(a) Paradigm 1: Creativeness as a Superhuman Quality

Up until fairly recently in human history, it was considered impossible for humans to be "creative". Instead, "creativity was attributed to a superhuman force; all novel ideas originated with the gods" (Sawyer 2012: 19), and the "creations" of seemingly "creative" human beings are to be traced back to them. In brief: superhuman entities such as the gods are the paradigm of creativeness. This is the Paradigm 1 view of who can be creative. The writings of Homer, Plato and Aristotle as well as the Bible – the "urtexts" of the pre-history of "creativity" – reflect this paradigm.

In ancient Greece, poets were not regarded as the originators of the epic poetry they voice, but merely as a medium through which deities – more specifically, the muses, the goddesses of poetic inspiration – speak.[4] The *Iliad* and the *Odyssey* mirror this view as they both commence with invocations of the muses. To quote the first verse from the *Odyssey*: "Sing to me of the man, Muse, the man of twists and turns" (Homer 1996: 1–3). Greek culture reveres poets such as Homer, Hesiod, or Archilochus because they are seen as communicators of the

[3] For comparative, cross-cultural perspectives on creativity, see, for example, Kakar and Blamberger (2015) and Weiner (2000).

[4] The following reflections on the idea of creativeness during antiquity draw on Blumenberg (2000), Curtius (2013), Mason (2003), McMahon (2013), Schmidinger (2008), and Weiner (2000).

divine, not because they create something new. So, in line with Paradigm 1, they were not regarded as "creative".

To make philosophical sense of the process of poetic composition, ancient philosophers introduced the notion of "divine inspiration". This idea goes back to Presocratic thinkers such as Empedocles and Democritus (Mackenzie 2021: 192–196) and was still upheld by Roman philosophers such as Cicero (1892: 105). However, its most influential expression can be found in Plato's early dialogue *Ion*. Here, Plato (or more precisely, the Platonic Socrates who is commonly considered Plato's mouthpiece) proclaims that the poets produce their works through "divine dispensation and possession" (1987: 536b 4–6) and that "the only thing they can compose properly is what the Muse impels them to" (1987: 534c 2). However, unlike most of his contemporaries, Plato holds that this is not a reason to revere poets as sources of wisdom, but a reason to mistrust them. Since poets are guided by a mysterious and irrational divine force, rather than by skill or rational principles, he contends that they do not know what they are talking about; therefore, they should not be treated as epistemic and moral authorities.

When one adopts Plato's broader dualistic metaphysical views on the nature of reality, the modern idea of *human* creativity appears even more preposterous. According to Plato's metaphysics, there are two spheres of being: the world of appearances which is perceptible to the senses, on the one hand, and a sphere of perfect, timeless, unchanging, mind-independent entities usually referred to as "Forms" or "Ideas", on the other. The physical objects that are accessible to sensuous experience are merely imperfect *imitations* of the perfect "Forms", dwelling in the Platonic heaven. On this metaphysical basis, it is impossible for poets to create something new. This is because Plato conceives of poetry in terms of *mimesis*, a concept that can be translated as "imitation", "copy", and "representation". Consequently, for Plato, poetry is an imitation of an imitation – and, as such, by its very nature not only illusive, but also mendacious and morally dubious so that Plato famously demands that poets ought to be expelled from an ideal city-state. Platonic metaphysics not only undermines the originality of

the poets – it leaves no space for genuine human creativeness whatsoever. Not only for poets, but for humanity in general, it is impossible to do something new as all putative innovations have already been pre-arranged in the immutable world of Ideas: "there are no unrealized Forms left over for the works of man" (Blumenberg 2000: 27, 29). So, human creativeness is nothing but an illusion.

Even though Plato's most famous student, Aristotle, challenged many of his master's views on the arts[5], he also denies poets a substantive role in the context of poetic creation. Contrary to his teacher, Aristotle holds that epic poetry and the arts more generally are not the product of divine inspiration, but are based on *téchne*, i.e., some kind of "craft", "skill", or "expert knowledge". Accordingly, Aristotle's *Poetics* (2013), his main work on poetic theory, aims to describe the most important rules of dramatic composition which are displayed by the supreme works of poetry. Like Plato, Aristotle conceives of art as mimetic. That, however, does not limit poetry to strictly copying something already existing. Rather, art, conceived as *téchne*, can – in line with Aristotle's teleological worldview – "complete[…] the things that nature is incapable of bringing to completion" (2018: II 8, 199a). This, however, drastically limits the creative powers of poets and humans more generally: "At the core of the Aristotelian concept of *techne* is the idea that man the maker cannot be assigned an *essential* [*wesentliche*] function. […] He brings to completion what nature *would* have brought to completion; the 'ought to be' is immanent in nature, not in him" (Blumenberg 2000: 32). Accordingly, due to the mimetic nature of their works, poets and artists more generally do not bring anything new into being. They are not more than "the extended arm of *physis*" (Kullmann 2010: 214), while the superhuman force of nature itself is the protagonist of artistic creation.

This shows that, while creative powers were linked to a superhuman force in antiquity, this force was not necessarily construed as the gods. The Roman thinker Lucretius (2007) goes

[5] The word "art" derives from "ars" which is the Latin word for *téchne*. Here, however, I use "art" in line with its modern meaning, as a concept that includes practices such as poetry, painting, sculpture, music, and dance.

one step further than Aristotle by explicitly denying that the gods created the cosmos (without denying their existence) and painting a materialist picture according to which nature itself is the supreme creative power. Even for Plato, the creative powers of the gods are limited. In *Timaeus* (1971), he argues that the highest god and fabricator of the cosmos, the so-called "demiurge", crafts the universe by using pre-existing materials which he shapes and arranges according to the plan set out by the eternal Ideas in the Platonic heaven. Thus, the demiurge does not *create* the world, but only *fashions* it according to the time-transcending Forms. So, the work of the artisan-like demiurge is ultimately also mimetic, imitative. In late antiquity, Neoplatonic philosophers painted a similar picture: according to Plotinus, for example, the universe does not come into being through a divine act of creation, but is an "emanation", or diffusion, from the transcendent source of all being, the "One".[6]

A resurrection and radicalisation of the idea that creative powers are restricted to a divine power is brought about with the advent of Christianity. In contrast to the Platonic demiurge and the Neoplatonic "One", the Christian God *is* a true "creator" as the opening words of the Bible make unmistakably clear: "In the beginning, God created the heavens and earth" (Genesis 1:1). Unlike Plato's demiurge, the God of the Bible does not merely fashion the world by re-arranging pre-existing materials; rather, he creates the world *ex nihilo*: "look at the heaven and the earth and see everything that is in them, and recognize that God did not make them out of things that existed" (2 Maccabees 7:28).

[6] An ancient account of poetic creation that challenges Paradigm 1, at least to some degree, is presented in a treatise from the first century AD titled *Perì Hýpsous* (*On the Sublime*). The anonymous author of this text is commonly referred to as "Longinus" or, in recognition of the uncertain authorship, "Pseudo-Longinus". According to Pseudo-Longinus, poetry and literature more generally are the product of the poet's mental capacities and "natural endowments" (Longinus 1890: VIII, 1). In other words: "sublimity", that is, literary greatness, "is, so to say, the image of greatness of soul" (Longinus 1890: IX, 2). So, in a sense, Pseudo-Longinus "humanises" poetic creation. However, Pseudo-Longinus's voice remained unheard by his contemporaries. There are no traces of it in the texts of other ancient writers. It was only in the late seventeenth century that *Perì Hýpsous* would attain widespread scholarly attention (Curtius 2013: 399). What is more, Pseudo-Longinus does not fully break free from Paradigm 1 thinking; accordingly, he writes that "the Sublime [i.e., literary greatness], lifts [the poet] near to the great spirit of the Deity" (Longinus 1890: XXXVI, 1f.). So, in *Perì Hýpsous*, creative powers in their highest form are still linked to the divine.

While the God of the Bible is an almighty God whose creative powers exceed those of the ancient gods, human making is regarded as derivative of God's infinite potency. This view was supported by the circumstance that the Hebrew Bible reserves the word *bara* (to create) almost exclusively to God; the Vulgate incorporates this linguistic differentiation to the effect that, up until the end of the Middle Ages, the capacity of "creatio" was almost exclusively linked to God. In line with that, Augustine insists that "[c]reatura non potest creare" (quoted from Lowinsky 1964: 477), that what has been created (by God) cannot create. Or as Thomas Aquinas puts it: "Solus deus creat" (1891: 220), only God creates. Therefore, humans are considered incapable of creation – to the extent that Augustine even denies them the ability to create children (Mason 1988: 709).

However, over the course of the Middle Ages, Paradigm 1 starts to exhibit fine cracks, mainly due to the emergence of Nominalist philosophy. This strand of medieval thought contemplates the status of so-called "universals", i.e., abstract entities such as general concepts that categorise individuals. In stark contrast to Platonic metaphysics, Ockham, the most influential exponent of Nominalism, assumes that "[n]o thing outside the soul is universal" (1994: 204). On this basis, abstract entities such as general concepts are not conceived of as pre-given, eternal, mind-independent entities, but mental or linguistic constructions, human creations, as it were. Nevertheless, up until late medieval times, the divine or superhuman had generally been considered the sole locus of creativeness, even if Nominalism had cast some critical light on this assumption. However, with the European Renaissance, the fine cracks Nominalism inflicted upon Paradigm 1 turned into more substantial ones.

(b) Paradigm 2: Creativeness as a Human Quality

While the verb *creare* had been reserved to God through the Middle Ages, a new understanding of the creative powers of humanity gradually breaks through with the emergence of the

European Renaissance.[7] The Florentine humanist Cristoforo Landino, for example, claims that the activity of the poets "comes very near to creating" (quoted from Abrams 1972: 273). Similarly, his contemporary Marsilio Ficino pronounces that "the power of the human person almost resembles that of God's nature" (quoted from Weiner 2000: 54). Nicholas of Cusa, one of the most influential representatives of the German Renaissance, even refers to humans as "creators". Reaffirming the Nominalist view that humans create the conceptual entities that allow them to categorise individuals, he writes: "just as God is the Creator of real beings and of natural forms, so man is the creator of conceptual beings and of artificial forms" (Cusa 1998: 794). The ideas voiced by Ficino, Landino, Cusa, and other Renaissance thinkers foreshadow a new paradigm of who can be creative, a paradigm in which the human replaces the divine and superhuman as the central instance of creativeness. In other words: "The Renaissance marked the beginning of the long 'transition' from God to human beings as the locus of creativity" (Glăveanu & Kaufman 2019: 11). However, the emergence of this new paradigm, Paradigm 2, was a gradual process. It took several centuries to fully unfold.

Even though Renaissance humanism paves the way for Paradigm 2, it still carries significant baggage from Paradigm 1. Early Renaissance thinkers such as Ficino, Landino, and Cusa were not only humanists, but also Neoplatonists. Therefore, they viewed art as an imitation of divine reality. Or as Michelangelo put it: "Good painting is nothing else but a copy of the perfections of God" (Buonarroti 1991: 77). Moreover, Renaissance thought is still in line with the Christian tradition insofar as human creations are considered derivative from God's infinite power. In the words of Landino: God is "the supreme poet, and the world is His poem" (quoted from Abrams 1972: 273). So, the creative power ascribed to poets, painters, sculptors, and architects is subordinate and inferior to God's limitless creative potency. As Cusa renders

[7] The following reflections on the idea of creativeness during the Renaissance draw on Mason (2003), McMahon (2013), Schmidinger (2008), and Weiner (2000).

it: the human intellect creates only “likenesses of the likenesses of the Divine Intellect” (1998: 794). Nonetheless, the terms “creation” and “humanity” are no longer mutually exclusive.

However, since creativeness is still connected to the divine, Cusa argues that humankind itself is divine, albeit in a secondary manner; in Cusa’s words: “man is a second god” (1998: 794). Similarly, during the Renaissance, it is common to praise accomplished artists as “second deit[ies]” (Scaligero 1905: 8), “mortal gods” (Vasari 1987: 284), or “creating gods” (Puttenham 2016: 94). So, in a sense, artists and humankind more generally must become divine in order to qualify as creative. (Albrecht Dürer’s iconic 1500 painting *Self-Portrait at Twenty-Eight* in which he depicted himself in striking resemblance to contemporary representations of Christ can be interpreted as a painterly expression of this idea.) Nonetheless, even if creativeness has not been fully cut off from the divine, it has become open to the human. So, in a way, Paradigm 1 and Paradigm 2 co-exist in the late Renaissance as this statement by Torquato Tasso illustrates: “[t]here are two creators”, namely “God and the poet” (quoted from Williams 1961: 22).

The gradual rise of Paradigm 2 goes hand in hand with a new *self*-understanding of the artist that emerges over the course of the Renaissance. Early Renaissance artists thought of themselves more as craftsmen, rather than “creators”. In line with that, they were organised in guilds like other craftspeople such as goldsmiths, masons, or carpenters. Towards the end of the Renaissance, however, things start to change. The artists emancipate themselves more and more from the other artisans and the guilds and establish their own academies such as the Florentine Academy of Art (Kristeller 2021: 176, 182). Furthermore, in the early sixteenth century, it becomes common practice for artists to sign their works (Weiner 2000: 57). On top of that, after the Republic of Venice had established the first codified patent system in 1474, patent laws were introduced all over Europe. As a patent is the intellectual property right granted to the human being that invented something, it implies that humans can invent things, i.e., that they can be creative.

Not only during the Renaissance, but also during the formative years of the European Enlightenment, Paradigms 1 and 2 compete with each other for dominance.[8] At the beginning of the so-called age of reason, the power of Paradigm 2 was still limited: Descartes, for example, still views God the "creator of all things that exist" (2017: 44) and Francis Bacon echoes the Renaissance idea of "mortal gods" (1964: 106). Over the course of the European Enlightenment, however, Paradigm 2 eventually takes the centre stage – especially with the emergence of the idea of *original genius* that plays a key role in eighteenth-century and nineteenth-century aesthetics. As Joseph Addison, one of the early proponents of this idea, puts it: geniuses are the "prodigies of mankind" (1804: 344). The original genius literally replaces God as the epitome of creativeness: "the genius assumed powers that once had been reserved exclusively for God and the gods" (McMahon 2013: xviii).

The idea of original genius is, to some extent, a counter-reaction to another influential cultural movement of the Enlightenment, namely Neoclassicism. Its agenda is summed up by Alexander Pope: "All that is left us is to recommend our productions by the imitation of the Ancients: […] in every age, the highest character for sense and learning has been obtained by those who have been most indebted to them" (1876: 148). Or as Winckelmann has it: "The only way for us to become great […] is the imitation of the ancients" (2013: 32). Unsurprisingly, the writings of ancient authorities such as Aristotle and Horace were used as guidelines for artistic production. This is also because Neoclassicism – in this respect, heavily influenced by Neoplatonism and Aristotle's take on *mimesis* – gravitated towards the believe that "[t]he human mind can only create imperfectly" (Batteux 2016: 5) – a remnant of Paradigm 1 thinking.

The idea of original genius – which refers both to a faculty for producing great works of art and to the artist exhibiting it – is in stark contrast to Neoclassicist principles. As Kant puts it: "genius is entirely opposed to the spirit of imitation" (2000: 187). Instead, the concept

[8] The following reflections on the idea of creativeness during the Enlightenment draw on Cassirer (2009), Kivy (2013), Mason (2003), McMahon (2013), Robertson (2020), Schmidinger (2008), and Weiner (2000).

of "genius" gets closely linked to the notion of "originality". Accordingly, Edward Young writes in his 1759 *Conjectures on Original Composition* that true genius produces "unprescribed beauties and unexampled excellence" (1971: 50). This implies that original genius can create new things – according to some voices, even *ex nihilo*. Thus, Shaftesbury writes that, rather than just imitating God's creation, a true genius "can imitate the Creator" (1999: 93), more specifically, his power to create from nothing. Similarly, according to Rousseau, "true creative genius […] makes something out of nothing" (1994: 81).

The emergence of original genius goes hand in hand with a new understanding of art: "No longer defined mainly by imitation, by *mimesis* of reality, art is understood now more in terms of creation" (Taylor 1991: 62). Besides that, a more mundane factor from which the emergence of the notion of "genius" and the underlying idea of human originality benefitted is the fact that, with the enactment of the Statute of Anne, Britain started to grant copyrights to writers in 1710. This had an important impact on literary production insofar as it incentivised distinctness and originality, since a lack thereof could lead to a denial of copyright: "As literature became a form of commerce writers had a new reason to display their uniqueness; for their own survival, in material terms, they needed to be in some way original" (Mason 2003: 108).

Whereas, in ancient Rome, "a genius was a guardian spirit, a god of one's birth who accompanied individuals throughout life, connecting them to the divine" (McMahon 2013: xiv), Enlightenment aesthetics conceives of genius as a faculty *inside* the artist. As Kant puts it: "[g]enius is […] an inborn productive faculty of the artist" (2000: 186). Even if "the author of a product that he owes to his genius does not know himself how the ideas for it come to him" and cannot "communicate to others precepts that would put them in a position to produce similar products" (Kant 2000: 187), genius lies *within* the artist. More specifically, genius is commonly treated as a "power of the mind" (Blair 1823: 26), which obviates any appeal to transcendent forces and divine inspiration to explain creative achievements. Instead, there is a

tendency to invoke the mental faculty of imagination, whose importance for poetic creation had already been highlighted by Sidney (2002: 85) and Shaftesbury (1999: 139), as the root of genius: "Imagination is the quality of all others most essentially requisite for the existence of genius" (Duff 1767: 6). In other words, "the force of imagination" (Saint-Lambert 2007) replaces imitation and inspiration as the source of creativeness and artistic excellence. In short: the realm of gods is ultimately superseded by the realm of human mind as the source of creativeness. Paradigm 2 finally triumphs.

Genius was not just the subject matter of scholarly debates, but quickly acquired the status of a cultural icon, especially due to the cultural impact of the Romantic movement. The Enlightenment was generally pervaded by a highly individualistic and exclusive conception of genius; in Kant's words: "genius is […] a rare phenomenon" (2000: 196). Others were more specific: according to Lavater, the "proportion of genius to the vulgar is like one to a million" (1790: 190), and, according to Jonathan Swift, there are not "more than five or six men of genius in an age" (quoted from Johnson 1846: 152). However, gradually, a more inclusive and egalitarian understanding of genius emerges. In 1742, Hume had already contended that, while history celebrates the genius of a few individuals, "a share of the same spirit and genius must be antecedently diffused throughout the people among whom they arise" (1993: 58). With the advent of the Romantic movement, this view becomes more widespread. Novalis, for instance, proclaims "the elevation of all human beings to genius" ["die Erhebung aller Menschen zu Genies" (1978: 681)]. The egalitarian re-conception of genius that took hold during the Romantic period can be seen as a radicalisation of Paradigm 2 thinking: *all* humans, not just a few exceptionally gifted ones, can be creative. As Friedrich Schlegel puts it: "You should demand genius from everyone" (1991: 2).

Paradigm 2 also becomes more inclusive in another respect. The historical narrative developed above focuses mainly on how poets and, to a lesser degree, painters and sculptors bring their works into being. In brief: the pre-history of creativity centres on *artistic*

creativeness. Accordingly, Shakespeare was widely celebrated as the purest embodiment of original genius (Addison 1804: 344f.; Herder 2008; Goethe 1985). Kant even went so far to claim that "genius[…] is a talent for art, not for science" (2000: 195). However, this view was far from representative for his times. During the Enlightenment, science joined the arts as a province of genius. In the words of Alexander Gerard: genius "qualifie[s] for making new discoveries in science, or for producing original works of art" (1774: 8). Especially Newton was widely admired as an example of supreme (scientific) genius. Hume even calls Newton "the greatest and rarest genius that ever rose for the adornment and instruction of the species" (1810: 279). When Ward coined the word "creativity" in 1875, his usage of the term did not restrict itself to the arts. As he introduces the term by speaking of "Shakespeare's […] poetic creativity" (1875: 506), he implies that there are also other (non-poetic) forms of creativity. So, the concept of "creativity" emerged as something than can occur in various domains (Weiner 2000: 8).

It took another 70 years from its coinage until the word "creativity" entered everyday language. Then, however, creativity soon turned into the subject of scientific, particularly psychological, research, especially after J. P. Guilford's highly influential 1950 American Psychological Association keynote address which laments the "neglect of this subject by psychologists" (1950: 445). On Guilford's definition, "creativity refers to the abilities that are most characteristic of creative people" (Guilford 1950: 444). Like Guilford, the emerging psychology of creativity treated its research object chiefly as a personality trait, a feature of persons. This perspective gave rise to two directions for researching creativity in persons: the first one deals with what is often dubbed "Big-C" creativity, that is creativity in extraordinarily creative persons (in geniuses, if you will); the second one deals with so-called "little-c" creativity, the creativity displayed by ordinary people.[9] As the latter notion demonstrates,

[9] These labels were introduced by Gardner (1993).

modern psychology rests on an inclusive notion of creativity, one which, as a report of the British *National Advisory Committee on Creative and Cultural Education* puts it, "recognises the potential for creative achievement in all fields of human activity; and the capacity for such achievements in the many and not the few" (1999: 30). By treating creativity as a personality trait of – highly original or ordinary – humans, the psychology of creativity has been erected on the basis of Paradigm 2 thinking.

From gods and superhuman entities to supreme artists and (with some delay) preeminent scientists to humanity more generally – the scope of who can be creative gradually expanded as this historical overview shows. Accordingly, some theorists speak of a normalisation (Reckwitz 2016: 200) and democratisation (Glăveanu & Kaufman 2019: 17) of creativity. Creativity is no longer seen as exceptional, but as something normal, achievable by every human. Yet, there are still biases when it comes to the attribution of creative abilities. According to a recent study (Proudfoot et al. 2015), men still tend to be perceived as more creative than women. Moreover, the notion of "genius" was used to support racist ideas, for example, in the work of Georges Cuvier who claimed that "[t]he white race[…] is also superior to others by its genius" (quoted from Baker 2006: 11). Even today, our practices of commemorating creative achievements in the arts and sciences still centre predominantly on white men which "carries the subliminal message that genius is male and caucasian" (MacLeod 2009: 573). So, despite the diagnosed democratization of creativity, it would be illusional to assume that the appreciation of creative achievements is always free from biases.

Moreover, ideas and assumptions linked to Paradigm 2 have not been uncontested. Think, for example, of Roland Barthes's pronouncement of the "Death of the Author" (1986: 49) and the Postmodernist vision of a state of "postcreativity" (Brown 1999: 423). Despite these efforts, Paradigm 2 is not dead. In fact, the provocative aura of these views stems from the circumstance that "creativity" is intimately connected to "the human". However, given the

impressive performance of state-of-the-art AI systems, the question arises if *only* humans (and maybe gods) can be creative or if machines also deserve the status of creative beings.

Even if it is a controversial matter if humanity has a monopoly on creativity, so to speak, humans still embody the paradigm of creativity. Accordingly, this inquiry (except for the Coda) focuses on the Paradigm 2 understanding of "creativity". It aims to obtain a better understanding of the current notion of "creativity" – and later examines if machines can be creative against the backdrop of this concept of "creativity". For that reason, the next two sections take a closer look at influential approaches to defining and theorising about "creativity" (in the Paradigm 2 sense) that play an important role in contemporary creativity research. We first assess widespread suggestions on how to define "creativity".

2.2 The Nature of Creativity: What is Creativity?

The almost unquestioned way of approaching this question in the scholarly literature consists in seeking individually necessary and jointly sufficient conditions for "creativity". Following this strategy, there is a strong tendency to define "creativity" as the capacity to produce things that are both new and valuable: "There is an emerging consensus that a product must meet two conditions in order to be creative. It must be *new*, of course, but […] also […] *of value*" (Paul & Kaufman 2014: 6). It has even been claimed that there seems to be "a general agreement that creativity involves the production of novel, useful products" (Mumford 2003: 110). According to this pervasive view, there are two necessary and jointly sufficient conditions for creativity, namely (i) a novelty, newness, or originality condition and (ii) a value or usefulness condition. Most creativity theorists agree that something counts as "creative" if and only if it satisfies these two conditions. This has been called the "standard two-part definition" (Gaut 2010: 1039).[10]

[10] Expressions of the standard two-part definition of creativity can also be found in Cropley and Cropley (2013: 5) and Kieran (2014: 128).

It is sometimes argued that Immanuel Kant anticipated the two-part definition of creativity (Guyer 2003: 116-37). In his reflection on genius in *Critique of the Power of Judgment*, Kant states that the genius necessarily requires originality, i.e., the ability to bring about novelties: "originality must be [genius's] primary characteristic" (2000: 186). Yet, Kant points out that it is possible to produce "original nonsense" (2000: 186), something that is novel, but without value. Kant's "original nonsense" argument is still a common point of reference in today's literature on creativity. One can, as Paul and Kaufman (2014: 6) contend, easily write down a random sequence of letters nobody else has ever put down on paper. Such a sequence is something new – yet, in most circumstances, it would lack any meaning and, thus, also any value. Since nonsense can be original, Kant insists that the works of a genius must "at the same time be models, i.e., exemplary" (Kant 2000: 186). Being a model or exemplary implies being valuable of some kind. So, Kant holds that the creations of a genius must be original *and* valuable. The advocates of the standard two-part definition, that is, the majority of creativity researchers today, agree that the same holds true for creativity – that "creativity" stands for "valuable novelty" (Csíkszentmihályi 2013: 51).

Even though the two-part conception is, more or less explicitly, the definitional basis of the bulk of today's creativity research, one also encounters a significant number of three-part definitions in the literature. Kaufman and Sternberg, for example, consider "new, good, and relevant" (2007: 55) to be essential properties. Yet, like this definition, most three-part conceptions can be reduced to the classical two-part definition. "Relevant", for example, seems to be just another expression for "valuable". However, there are also some genuine three-part or n-part definitions that invoke further candidates as necessary conditions of "creativity" that cannot be reduced to "novelty" or "value". So, the two-part definition is not uncontested.

In the following, we will first discuss to what extent "novelty" and "value" are necessary for "creativity". After that, we will attend other concepts that are introduced by influential

theorists to complement (or replace a part of) the two-part definition and scrutinise to what extent these further candidates can serve as necessary conditions for "creativity".

(a) Novelty Condition

The two-part as well as virtually all three-part or n-part definitions take for granted that creativity entails novelty. However, the main issue with the novelty condition is that it is unclear what counts as "novel", "new", or "original". Margaret Boden who deems novelty an essential feature of "creativity" distinguishes between what she calls "P-creativity" and "H-creativity". Something is P-creative (that is, "personally" or "psychologically" creative) if it is "*new to the person who comes up with it*" (Boden 2004: 2). Yet, if an idea or artefact is H-creative (that is, "historically" creative), "that means that (so far as we know) no one else has had it before: it has arisen for the first time in human history" (Boden 2004: 2). (Since a creative achievement that is historically unprecedented must be new to the creator, H-creativity necessarily implies P-creativity.) One can also put it like this: H-creativity generates H-novelty (historical novelty), while P-creativity only generates P-novelty (personal novelty).

However, there are cases of creativity that are neither H- nor P-novel. Modernist luminaries such as Marcel Proust, Virginia Woolf, James Joyce, and Alfred Döblin are globally admired for presenting "streams of consciousness" in their texts. It is widely accepted that all these writers did something creative when they made use of this narrative device. However, it is common knowledge amongst literary scholars that this technique had been employed before it made its most celebrated appearance in modern masterpieces such as *Ulysses* or *To the Lighthouse*. In Édouard Dujardin's 1888 novel *Les Lauriers sont Coupés*, one can already spot this narrative device in its full-fledged form. Indeed, William James coined the term "stream of consciousness" to describe the literary mechanisms he encountered in Dujardin's novel. Nabokov even contended that Tolstoy's *Anna Karenina* showcases the first stream of

consciousness in literary history (1981: 183). Others might go further and argue that the technique can even be traced back to Edgar Allan Poe or Laurence Sterne.

This shows that, whoever it was who invented the technique, it certainly was not Proust, Woolf, Joyce, or Döblin. Consequently, they did not do something H-novel (or H-creative) by using the stream-of-consciousness technique. Furthermore, at least some of them did not do something P-creative either. Joyce, for example, made no secret of the fact that he did not develop this technique on his own (in an act of P-creativity), but encountered it in Dujardin's work (Beja 1992: 66). Thus, his use of this literary method was not only not H-novel, but also not P-novel. Nevertheless, there is little doubt that the display of the stream-of-consciousness technique in the last chapter of *Ulysses* represents a creative achievement.

Can we, despite the absence of H- and P-novelty, still say that Joyce did something new by using the stream-of-consciousness technique? The defenders of the novelty condition can emphasise that "novelty" comes in degrees. Accordingly, they can contend that, even if Joyce's use of the stream-of-consciousness method might be less novel than Dujardin's, we should still attribute a significant amount of novelty to it. For, when Joyce wrote *Ulysses*, this literary technique was still relatively unknown and, thus, perceived as new.

The graduated nature of "novelty" raises the question of where we set the threshold for novelty when making judgements about creativity. From a certain perspective, every product of human activity is new. From a different perspective, there is, as the Book of Ecclesiastes famously puts it, "no new thing under the sun". Most people will probably not take such extreme conceptions of "novelty" as the basis of their assessments of the novelty of creative achievements. Nonetheless, there is still significant disagreement as to what qualifies as "novel" when assessing the novelty of creative achievements.

While many people would probably accept that Joyce's use of the stream-of-consciousness technique exhibits some degree of novelty, there are cases that are more contentious. How about using the stream-of-consciousness method in 1985 (i.e., more than 60

years after the publication of *Ulysses*), for example? Or how about *Grimms' Fairy Tales*? The Brothers Grimm did not invent, but only collected, wrote down, and adapted the stories they are known for. How one answers these questions depends on one's conception of "novelty". While some people would probably ascribe at least some degree of novelty or originality to these examples, others would refuse to do so. In the face of these disagreements, one best accepts the controversial status of the novelty or originality of these examples. More technically put: since we take these disagreements seriously, we treat such contentious examples as borderline cases when it comes to the ascription of novelty. Hence, there are creative achievements whose novel and original status is disputable.

While the creativity of the Brothers Grimm and "their" fairy tales seems to be widely unchallenged, it might be more controversial to claim that using the stream-of-consciousness method in 1985 really qualifies as "creative". And calling a masterful replication of a famous artwork "creative" seems to be even more contentious. Disagreement regarding the creative status of certain things is often due to the contentiousness of their novelty or originality. Consequently, *some* of the cases that are borderline cases as far as the ascription of novelty is concerned are also borderline cases when it comes to the ascription of creativity.

(b) Value Condition

While the novelty condition of the standard two-part definition is barely questioned in the literature, the value condition has recently been exposed to some criticism. To challenge the value condition, several critics point to examples of what one might call malevolent creativity. It is, for instance, possible to be creative in carrying out a devastating terror attack (Cropley et al. 2008; Beaney 2005: 190f.), devising new execution and torture methods (Hills & Bird 2019:

701; McLaren 1999; Rogers 1970: 139), or finding new ways to avoid taxes (Hills & Bird 2019: 701). These things, the argument goes, are still creative, but have no (positive) value.[11]

"Value" is a difficult and philosophically charged concept. But it seems right that many people would agree that these things are not valuable. Yet, this assessment seems to be due to an overly "moralistic" conception of "value". If you have Utilitarian intuitions, things appear valuable if they lead to good and desirable consequences, if they eventually "tend to promote happiness" (Mill 1998: 122). On the other hand, if you lean towards a more Kantian position according to which there is nothing "that could be taken to be good without limitation, except a good will" (Kant 2012: 9), the value of an action is tied to the creator's will and intentions. Alternatively, if you gravitate towards virtue ethics, value is linked to virtues.[12] Whatever your ethical outlook, it is hard to justify the moral value of, say, creatively conducting a horrible terror attack.

However, such a "moralistic" conception of "value" is too narrow. So, a common argumentative move consists in introducing the notion of something being *good of its kind*. Accordingly, the "creative terrorist is a good terrorist, in the sense of being good as a terrorist" (Gaut 2018: 128). Some produced kinds tend to be morally valuable (such as new medicines, scientific theories, or artworks), while others are not (such as new torture techniques or tax-avoidance strategies). That, however, does not mean that the latter lack *any* value. For, even if something is morally bad, it can still be good of its kind.

Appeals to kind-specific value, however, have engendered criticism. Alison Hills and Alexander Bird, for example, rightly observe that things normally belong to more than one kind. So, it seems more accurate to say that "creatively produced ideas are good *qua some relevant kind*" (Hills & Bird 2019: 702; emphasis added). But Hills and Bird reject this approach as well

[11] Novitz (1999: 78; 2003: 186f.) rejects the idea that cases of malevolent creativity are truly creative. For a convincing critique of Novitz's argument, see Livingston (2018) and Grant (2012).
[12] Some creativity researchers conceive of creativity as a virtue; see, e.g., Audi (2018), Kieran (2014), or Zagzebski (1996: 167).

by contending that there are examples of creativity that are not good or valuable qua any relevant kind. As alleged cases of value-free creativity, they invoke, for example, Percival Lowell's speculations about the existence of canals on Mars (2019: 702), Nikola Tesla's misguided theories about spacetime and the electromagnetic field, or William Herschel's view that the Sun is habitable (2019: 705; 2018: 98, 101). However, defenders of the value condition can respond that Hills and Bird wrongly presuppose that these things are devoid of any value and point to the circumstance that their examples exhibit a high degree of intellectual agility and resourcefulness. So, they might argue that, even if Lowell's, Tesla's, Herschel's above-mentioned ideas did not prove scientifically valuable, they are valuable as examples of intellectual agility and resourcefulness. To support that view, they might point to the circumstance that, while these ideas seem bizarre today, it was (as in the case of most ambitious intellectual endeavours) not clear that they would eventually turn out to be scientifically unsuccessful. However, the critics of the value condition can, of course, question the *relevance* of this kind of value (as intellectual agility and resourcefulness can also lead us astray).[13]

Hills and Bird do not draw on any philosophical theory of value to support their assumption that their examples showcase value-free creativity. Rather, they simply claim that one would not ascribe any value to these cases and take it for granted that this assessment is in line with the intuitions of the reader. In doing so, they disregard the fact that there is often considerable disagreement about what is valuable (or relevantly valuable) and what is not. It may seem tempting to resolve this disagreement by introducing a theory of value that tells us if there is any relevant value at play in the examples presented by Hills and Bird. Yet, there is no

[13] Apart from that, Hills and Bird argue that we do not need to know the value of something to know if it is creative: we do, for example, "not need to know whether Leonardo's designs [of flying vehicles] stood any chance of working […] to judge that these ideas manifested his creativity" (2019: 705). Again, the defenders of the value condition can respond that Hills and Bird miss that we value Leonardo's sketches as instances of intellectual agility and resourcefulness. Besides that, Hills and Bird also point to the circumstance that some individuals such as Tesla and Herschel have created both good and bad ideas and argue that it is psychologically "not plausible that one set of dispositions (creativity) produced the good ideas of Tesla and Herschel and quite another generated the bad ones" (2019: 705). Here, Hills and Bird reduce "creativity" to idea generation; this view will be challenged in section 2.3.

indisputable theory of value either. If a theory of value tells us that, say, Lowell's speculations about Martian canals exhibit no relevant value, those who do not share this intuition can disagree with this theory of value by drawing on their intuitions. A theory of value everybody can agree on is not available. (See the abounding disagreements in meta-ethics about the nature of moral value.) So, appealing to such a theory will not solve the problem.

As there is no available strategy to reconcile the disagreement about whether the above-mentioned ideas of Lowell, Tesla, and Herschel are valuable in any relevant sense, we do best to proceed by simply accepting these disagreements. Accordingly, we should accept these examples as contentious cases of value-free creativity. To put it differently, if we take both sides of the disagreement seriously, there is nothing more that we can do than regard these examples as borderline cases of relevantly valuable (or, if viewed from the opposite angle, value-free) ideas.

Hills and Bird not only posit that the ideas of Lowell, Tesla, and Herschel do not manifest any relevant value – they also posit that that they are examples of creativity. However, this second assumption is also more contentious than Hills and Bird acknowledge. At least some people who agree with them that the examples they provide do not exhibit any relevant value might, at the same time, disagree with them about whether these cases exhibit creativity. (It is not for nothing that the majority of creativity researchers take for granted that value is an essential feature of creativity.) So, if something proves to be valueless, this might be a reason for many people to deny that this thing can be creative at the same time. As stated above, there are borderline cases of creativity whose controversial creative status is due to the controversial status of their originality. Likewise, there are also borderline cases of creativity whose controversial creative status is due to the controversial status of their value. As noted, there seems to be a lot of potential for significant disagreement about whether Hills and Bird alleged examples of valueless creativity really qualify as "creative". Thus, we should regard those specific cases that trigger such disagreement also as borderline cases of *creativity*.

(c) Imagination Condition

Since Hills and Bird reject the view that creativity needs to be valuable, they argue that something else ought to take the place usually occupied by value as a supplement to originality when it comes to defining "creativity": "Rather than value, we propose that the imagination is essential to creativity" (2018: 105; also see 2019: 695-700). And they are not alone in treating imagination as a necessary condition of creativity: for Langland-Hassan, "there are no creative acts in which the creator's imaginativeness played no role" (2020: 263). Similarly, for Stokes, "imagination is important for even the most minimally creative thought" (2014: 158). Audi is even more straightforward, calling "[i]magination […] the chief constituent in creativity" (2018: 27).

This raises the question of how to conceive of "imagination". Many philosophers stress the slipperiness of this concept (Stevenson 2003; Walton 1990: 19). Accordingly, Hills and Bird cautiously describe imagination as "an ability to produce a particular type of mental representation" (2019: 696), without elaborating on what is particular about that type. They maintain that the imagination can generate new mental representations willingly and unwillingly as well as consciously and unconsciously. Moreover, they assert that imagination can, but does not need to, be accompanied by mental imagery (Hills & Bird 2019: 696). As imagination enables us to produce the relevant kind of mental representation, it allows us to search through a space of possibilities which represents, according to Hills and Bird, the source of creativity.

However, the idea that creativity presupposes imagination has been criticised. Gaut, for example, contends that there are examples of creativity that do not involve imagination. More specifically, he points to sudden *aha* moments or insights that occur to the creative individual while sleeping. One of his examples is Bertrand Russell who, in spite of intense intellectual efforts, often failed to find a solution to an issue he was tackling in *Principia Mathematica*, only to know the answer when he woke up the next morning. According to Gaut, "Russell went

from not knowing the answer to knowing the answer, without it seems any imaginative act on his part" (2003: 155). Gaut also invokes Friedrich von Kekulé who reported that, after extended struggles to identify the structure of the benzene molecule, the solution – that is, the idea that they are ring-shaped – occurred to him when he had a dream about a snake biting its own tail: "This example does involve imagery, but being dream-imagery, […] it may well not have involved imagination" (Gaut 2003: 155). Gaut does not mention these examples, but there are also several composers, including Stravinsky, Ravel, or Berlioz, who reported that beautiful melodies came to them in their dreams (Sacks 2007: 282f.).

Gaut's argument rests on the assumption that dreaming is conceptually distinct from imagining; so, defenders of the imagination condition could respond by contending that the underlying conception of "imagination" is too narrow. Carroll, for example, claims that "it is not clear that exercising the imagination is one and only one kind of mental activity. Rather it seems to be something more like a genus with several species vaguely united in that these species each go beyond what is given or has been given to the perception as fact in either the past or in the here and now" (2014: 63). Many dreams also go beyond current or past perceptions. So, if transcending what is currently perceived or has been perceived in the past is the sole requirement for some mental activity to qualify as an exercise of "imagination", then at least some dreams represent exercises of imagination. Dreams are normally not under the control of the dreamer's will; yet, as noted previously, the same pertains to imagination. On this basis, one could argue that, when Kekulé, Russell, Stravinsky, etc. had their scientifically, philosophically, or artistically insightful dreams, they were (unwillingly) using their imagination. Treating these kinds of dreams as instances of imagination is also compatible with Hills's and Bird's idea that imagination allows us to search through a space of possibilities.

Amy Kind recently presented another putative counterexample to the claim that creativity necessitates imagination: she asks us to imagine a pupil who, while doing her maths homework, "took an unusual pathway through the various mathematical rules that had been

previously taught to her" (2022: 36). She got the right answer, but her teacher had never seen her approach been taken before. Kind remarks that "we might note that the steps that she took to work her way through the problem seemed completely natural and even automatic to her" (2022: 36). In that case, Kind concludes, the pupil was creative without using her imagination. However, if one agrees with Carroll that the only hard requirement for some mental activity to qualify as imagination is that it transcends what is or has been given to perception, and accepts that imagination can be exercised unconsciously, one might draw the conclusion that the pupil *was* making use of her imagination. For the mathematical steps she mentally passed through went beyond the examples she has been exposed to before during the maths lesson. Likewise, if one concurs with Hills and Bird in conceiving of imagination as the mental search through a space of possibilities, one could claim that imagination is involved in Kind's example.

However, that does not mean that there are no cases of creativity without imagination. Consider a jazz saxophonist who performs an improvisation that features a creative tonal pattern. The advocates of the imagination condition may argue that the saxophonist is doing so by using her imagination, i.e., that she generates auditory mental representations that go beyond her past and current auditory perceptions. Yet, this account of what is going on here is mistaken. For it implies that the saxophonist, first, imagines a tonal pattern and, then, plays it. In doing so, it artificially separates performing the improvised tonal pattern from composing it by means of imagination. Yet, in fact, "[i]mprovising is simultaneously composing and performing" (Asma 2017: 48). Moreover, this mentalistic account conceives of the saxophonist's improvised performance in a disembodied manner. It envisions musical improvisations chiefly as mental achievements and neglects the importance of the performer's "muscle memory".

It might be the case that musicians who participate in an improvisation sometimes imagine how a tonal sequence might sound before playing it. But let us assume that the protagonist of our example performs her saxophone improvisation without imagining the tonal pattern in advance, but simply by relying on her muscle memory. This scenario does not require

her to have no mental content or representations at all, while performing her improvisation. She might mentally represent the auditory pattern of her improvisation in her mind as she is perceiving it. Yet, this would not count as an act of imagination since the auditory content or representations in her mind do not go beyond what she is currently perceiving. So, the mental representations accompanying her performance were not generated by her imagination. Thus, the musical patterns of the saxophonist's improvisation are not the product of her imagination – even though they exhibit creativity.

Since there is, at least in the creativity-related debate, strong agreement that imagination is some kind of *mental* or *internal* faculty, this counterexample affects practically all conceptions of "imagination" propounded by the defenders of the imagination condition (including the relatively undemanding ones that have been proposed by Carroll or Hills and Bird). An exception might be the conception presented by Robert Audi according to which "imagination […] is largely the capacity to produce—often initially in the mental realm—new things" (2018: 28). Audi does not really elaborate on this, but the "often initially" in the parenthesis suggests that, in his view, imagining does not need to be a mental activity. On this basis, one might claim that a jazz improvisation qualifies as an imaginative exercise. However, if one drops the mental activity component of Audi's conception, imagination is solely defined by being the capacity to produce new things. In other words: the concept of "imagination" becomes so semantically diluted that it becomes increasingly unclear what separates imagination from creativity. Saying that the capacity to produce new things is essential for the capacity to produce (borderline cases notwithstanding) new and valuable things is tautological. Hence, unless one dilutes the notion of "imagination" to such an extent that the assertion "creativity requires imagination" is a tautology, the claim that imagination is essential for creativity cannot be maintained.[14]

[14] To de-mentalise "imagination", it might also be appealing to conceive of this capacity in terms of behavioural criteria, even if this approach is practically non-existent in current definitional debates about creativity. According

(d) Improvisation Condition

There is one account of "imagination" in the philosophical literature that might be able to cope with our counterexample of the improvising saxophonist. This account has been developed by Stephen Asma and maintains that imagination and improvisation are not only compatible, but that the latter helps us to cast light on the workings of the former: "Improvisation, in my account, will be the main *activity*, method, or operation of the imaginative faculty" (2017: 4). This claim needs be understood against the backdrop of Asma's critique of the status quo of the philosophy of imagination, the circumstance that virtually all "philosophers characterize imagination as a kind of *cognition*" (2017: 4) or mental activity. Conceiving of imagination in terms of improvisation allows Asma to de-mentalise imagination, as it were. In doing so, his take on "imagination" radically diverges from the above-mentioned mentalistic conceptions that are proposed or adopted by the main advocates of the imagination condition.

This leads to the following question: should we consider improvisation an essential feature of creativity? More specifically: does creativity require improvisation, construed as an embodied phenomenon that does not solely take place inside the mind? Asma states that "[t]he improvising imagination draws on internal resources (i.e., thoughts, feelings, behaviors) and environmental resources (i.e., this tool, this pigment, this rock) in the service of various end goals" (2017: 4). In other words, improvisation bridges the inner and the outer.[15]

Invoking improvisation, thus construed, as a condition for creativity, however, also proves problematic. For there are examples of creativity that do not rely on improvisation as an activity that mediates between the inner and the outer, namely: creative dreams à la Kekulé,

to this approach, we ascribe "imagination" to persons based on their behaviour in specific contexts, i.e., when they behave imaginatively. Yet, "imagination", thus conceived, still does not represent a necessary condition for "creativity". For we can imagine circumstances in which one (or, at least, jazz aficionados) would not ascribe "imaginativeness" to the improvising saxophonist and her creative performance, e.g., when the tonal structure follows a standard jazz pattern. (In section 2.3, we will say more about the creativity of this kind of musical improvisation.)

[15] Since Asma lists behaviours as an internal resource, the improvised performance of the above-mentioned jazz saxophonist can still be understood as something that bridges the inner and the outer on his account, even though her musical performance is not the product of her imagination, mentalistically construed. For the saxophonist relies on her muscle memory – and the latter can be understood in terms of behaviour.

Stravinsky, and Ravel. It remains unclear how creative dreams draw on environmental resources (as conceived by Asma) in any significant sense.[16] Therefore, it seems more plausible to conceive of the creative dreams of Kekulé, Stravinsky, etc. as the product of unwilling imagination (understood as an internal and mental activity), rather than the outcome of improvisation (understood as an intermediary between inner and outer). Thus, we can conclude that improvisation (in this sense) is not essential for creativity.[17]

(e) Intrinsic Motivation Condition

Some researchers argue that creativity also requires motivation – and not just any kind of motivation, but *intrinsic* motivation. For Teresa Amabile, intrinsic motivation is part of "a set of necessary and sufficient components of creativity" (1983: 357). Amabile writes: "Intrinsic task motivation is passion: the motivation to undertake a task or solve a problem because it is interesting, involving, personally challenging, or satisfying—rather than undertaking it out of the extrinsic motivation arising from contracted-for rewards, surveillance, competition, evaluation, or requirements to do something in a certain way" (2013: 135f.). Hills and Bird also regard motivation as a condition for creativity; even though they speak of "motivation" only, it becomes apparent that what they mean is intrinsic motivation, as they write that creative people have an "urge to create" (2019: 700).

[16] Dreams might draw on memories of external phenomena. Yet, such memories still represent internal resources. Even if one accepts that there is external memory on the basis of the extended mind hypothesis (Clark & Chalmers 1998), external memory stores such as notebooks or PCs are not accessible during sleep.

[17] Asma himself briefly touches on dreams and remarks that "[d]reams are improvisations in the sense that they are autonomous, uncontrolled narratives with loose cause-and-effect sequencing" (2017: 110). These narratives are in the mind of the dreamer. This seems to contradict Asma's above-mentioned remark that improvisation draws on environmental resources, his view that "[i]mprovisation[…] is the adaptive meeting place between the organism and the environment" (2017: 4). (To be fair, even though Asma considers improvising imagination "the enigmatic engine of human creativity" (2017: 1), it needs to be mentioned that he is not primarily dealing with the issue of how to define "creativity", but of how to explain imagination, without reducing it to cognition.) Irrespective of whether creative dreams qualify as genuine improvisations on the basis of Asma's conception, we do not claim here that creative dreams cannot *in some sense* be described as "improvisations". Rather, the point we are making is that it is mistaken to classify such dreams as improvisations if "improvisation" stands for an embodied activity that draws not just on internal resources, but also external ones.

However, there is no reason to assume that creativity necessitates intrinsic motivation. Think, for example, about Scheherazade from *One Thousand and One Nights* and the 1,001 tales she tells King Shahryar to prevent being killed. Or think of a person that hates playing chess but is forced to play a game and makes highly original and effective moves. Many people would still call the chess player and his moves as well as Scheherazade and her stories "creative" – despite the absence of intrinsic motivation.

(f) Agency Condition

Some theorists argue that creativity necessarily requires agency: "the right kind of process for creative production is one that involves, in some non-trivial way, the agency of the creator" (Paul & Stokes 2018: 197).[18] Or as Berys Gaut, the main proponent of this view, puts it: creativity is an "exercise of agency" (2010: 1041); accordingly, "only agents and their products can be creative" (2018: 130). Gaut suggests a three-part definition according to which novelty, value, and agency are necessary and sufficient conditions for creativity (2010: 1040f.). We have seen that the first two conditions are debatable due to the existence of borderline cases. Yet, agency is still a contender for an undisputable *conditio sine qua non* for creativity.

Gaut defends the agency condition by claiming that "[b]eing creative is incompatible with doing something purely by luck" (Gaut 2010: 1040; also see 2018: 130). He presents the following scenario: you accidentally spill paints onto a canvas, and, by sheer luck, they form a highly aesthetic and original painting. If that happened, Gaut writes, "[y]ou would not be creative, since the painting was purely the product of luck" (2010: 1040). Therefore, he argues that "merely being produced by an agent—being the product of an agent-involving event—cannot suffice" (Gaut 2018: 130). On that basis, Gaut insists that creativity does not only require agency, but *intentional* agency: "To be creative, then, an agent's actions have to have some sort

[18] For the view that creativity is a feature of agents also see Carruthers (2006: 277–334) and (Stokes 2008; 2011; 2014).

of intentional connection with the kind of values for which he is being credited as being creative" (Gaut 2018: 130). In brief: for Gaut, a product or agent only qualifies as creative if the latter *intends* all or, at least, some (2018: 132) of the values exhibited by the former. Paul and Stokes point in the same direction when they write that "creativity always involves intentional action" and that "x is creative only if x is the non-accidental result of agency" (2018: 197).

In his latest publications on the subject, Gaut emphasises that, in order to be able to intentionally produce values, an agent needs to know how to produce these values: "what is required to be creative is that one exercise[s] *knowledge* of *how* to produce something with the relevant values, that is, one must *understand* how to produce them" (2018: 131). To sum up: Gaut holds that, in order to be creative, an agent must intend to create a product that features certain values and must have the relevant knowledge and understanding to do so. By invoking these conditions, he tries to make sure that creativity is not ascribed to things that are merely the result of a lucky accident or to the people who benefit from such lucky incidents.

Gaut develops his agency-based account solely with reference to artistic examples, i.e., products featuring (or lacking) artistic value. Many of the intuitions these examples trigger seem to depend upon the Idealist preconception that true artists autonomously materialise their vision in their works and have everything under control while doing so. However, the history of science presents a very different picture of how scientific discoveries often come about: some of the greatest scientific achievements are the result of luck and blind coincidences – nonetheless, they are celebrated creative achievements. The discovery of penicillin is arguably the most famous example: when Alexander Fleming returned from his summer holiday in September 1928, he noticed that a greenish mould had contaminated the colonies of staphylococci bacteria he had prepared for future experiments. When he examined the contaminated petri dishes under the microscope, he was astonished to see that the mould, called

Penicillium notatum, had stopped the growth of the staphylococci (Markel 2020: 287). By accident, Fleming had stumbled across penicillin, the first antibiotic in medical history.

Later, Fleming honestly admitted: "When I woke up just after dawn on September 28, 1928, I certainly didn't plan to revolutionize all medicine by discovering the world's first antibiotic, or bacteria killer. But I guess that was exactly what I did" (quoted from Markel 2020: 287). This shows that he did not have the intention to develop penicillin. Let alone did he know how to produce a bacteria-killing substance (or the values associated with it) before he entered the lab on 28 September 1928. It was due to a lucky accident that he discovered penicillin. Nevertheless, Fleming's discovery is commonly considered one of the greatest triumphs of twentieth-century medicine and, as such, a creative accomplishment.[19]

It has been demonstrated that penicillin was discovered even though its discoverer did not intend to discover it (as he did not search for a substance to treat deadly bacterial infections) and had no clue about how to develop such an antibiotic (if he had aimed to discover or invent one).[20] Yet, one might argue that this only shows that creativity does not require *intentional* agency. That, however, does not mean that there was no agency whatsoever at play here. For Fleming is an agent – and he still plays an important part in the story of the discovery of penicillin. Even though he did not intend to develop an antibiotic when he went to his laboratory on 28 September 1928, one can argue that he displayed "creative insight in discerning the

[19] Gaut does not hold that creative work needs to be totally free from chance. However, illustrating this by pointing to the artistic work of Francis Bacon, Gaut argues that, if chance is involved, it is employed "as part of an intentional process that [is] sensitive to painterly values" (2018: 132). In brief, Francis Bacon used chance or randomness *intentionally*. Yet, it would be a mischaracterisation to say that Fleming intentionally used chance or randomness as he did not have the intention to develop an antibiotic. So, a more adequate way of putting it is to say that a lucky accident happened to him. In line with that, Fleming is said to have said that "[o]ne sometimes finds what one is not looking for" (quoted from Cropley 2019: 119).

[20] Instead of know-how and understanding, Gaut occasionally also speaks of "skill" (2018:131), but, again, mainly with regard to artistic production. Yet, it seems fair to say that, without the lucky accident, Fleming would have lacked the skill to produce an antibiotic like penicillin in the same sense that, say, Michelangelo had the skill to produce a large wall fresco. When the former woke up on 28 September 1928, he was unable to produce a cure for bacterial infections. In contrast, the day that Michelangelo started painting the fresco of the Sistine Chapel, he knew what he had to do to achieve his goal. Otherwise, the Vatican would not have entrusted him with the task of painting the chapel's ceiling in the first place.

significance of the accident[…] that occur[ed]" (Martin 2007: 52f.).[21] Even if Fleming was not actively searching for a cure for deadly bacterial infections, he recognised that the mould in the petri dish offered an opportunity to find one.

Critics of the agency condition can object that this does not qualify as genuine agency. Some influential figures in the philosophy of action, most notably Anscombe (1957) and Davidson (1963), explain agency in terms of intentionality. In their view, the latter is necessary for the former. So, according to this conception, agency that lacks intentionality does not count as genuine agency. (To avoid misunderstandings: in the current section, "intentionality" is used in its everyday sense, referring to having an intention, and not in the technical sense that pervades contemporary philosophy of mind, referring to the representational powers of the mind.) If this is true, one can argue that simply discerning the significance of a lucky incident does not always manifest genuine agency. For many proponents of the intention-based account hold that agency requires that an action is causally initiated by the relevant intentions of the agent (Brand 1984; Bratman 1987; Mele 1992, 2003). On this basis, one can contend that Fleming did not *act* by discerning the significance of his lucky accident as he did not intend to find a treatment for bacterial infections and, thus, lacked the relevant intentions that are required to causally initiate an action.

However, even though intention-based conceptions are an important point of reference in current philosophical debates about the nature of agency, they are far from universally accepted and raise various philosophical issues (for an overview of some of these problems, see Schlosser 2019). If one denies the centrality of intentionality for agency, as many philosophers do, one is more likely to accept that the discovery of penicillin involved agency.

[21] To do so, Fleming required some know-how that allowed him to see that the growth of the bacteria in the petri dishes had been inhibited and to recognise the kind of mould responsible for it. Yet, the know-how for producing an antibiotic was only acquired by him over the course of further investigations that were triggered by the lucky incident.

However, even if one decouples agency from intentionality, it is still controversial whether creativity *always* involves agency. If the underlying conception of agency is vague and undemanding enough, it seems almost always possible to see some agency at play. Yet, critics of the agency condition can still challenge the significance of the kind of agency that is claimed to be involved. To illustrate this, consider the following scenario: a painter paints a painting on a canvas. Whenever she uses a new colour, she rubs the paintbrush against a piece of paper to get rid of the remining paint that is still on the bristles. The next day, a courier comes to the atelier of the painter to collect the artworks for an upcoming exhibition. The courier who is not an art expert inadvertently also takes the piece of paper that was used to clean the brush with him. Due to some logistical confusions, this piece of paper which was (obviously) not intended to be part of the exhibition is also displayed – and celebrated by art critics and the audience as a creative achievement. Let us further assume that the artist would not have been able to paint this accidental artwork consciously and willingly, as it were, as she lacked the skill to paint in the style it exhibits.

There is agency involved in this scenario insofar as there are agents involved, namely the painter who unintentionally and unconsciously "painted" the picture, the courier, the people in charge of the logistics and organisation of the exhibition, and, arguably, the audience and art critics who are fond of the accidental artwork. As noted, one can question the significance of the agency involved. In the penicillin example, Fleming recognised what was going on in the petri dish and noticed its significance which required a considerable amount of domain-specific know-how and perspicacity. So, one can argue he played a somewhat *active* role in the discovery of penicillin. Yet, in the example at hand, there is no agent that plays a similarly active role. Thus, critics of the agency condition can demand that an agent or a group of agents needs to play an active, rather than a merely passive, part in the production of a creative achievement. On this basis, they can contend that there are cases of creativity that lack active agency – and are, therefore, devoid of agency in any significant sense. (Of course, what counts

as an active agential contribution is also open to debate. Accordingly, some might even refuse to consider the agency exerted by Fleming to be active enough.) It is unlikely that defenders of the agency condition will accept this point of criticism. They might, for example, simply reject the claim that passive agency is an improper form of agency. Or they might deny that the accidental artwork qualifies as truly "creative".

In brief, whether creativity requires agency depends on one's conscious or unconscious (pre-)conception of agency. Analogously, intuitions about whether agency is (in any significant sense) involved in certain cases of creativity vary considerably. These opposing (pre-)conceptions and intuitions are reflected by the disputes in today's philosophy of action about how to conceive of agency (Schlosser 2019). In the face of these disagreements, we proceed by assuming that there are borderline cases when it comes to the question of whether agency is involved, for example, the discovery of penicillin or the production of the accidental artwork.[22] Moreover, as noted, there might be significant disagreement about whether some of these cases like the accidental artwork case really involve creativity. In other words, some of the cases that are borderline cases when it comes to the involvement of agency are also borderline cases of creativity.

(g) Spontaneity Condition

Even if he highlights the importance of intentional agency, Gaut maintains that there are limits to intentional control. Therefore, he propounds what he calls the "*Ignorance Principle*": "If someone is creative in producing some item, she cannot know in advance of being creative precisely both the end at which she is aiming and the means to achieve it" (Gaut 2018: 134). It follows from this principle that "all creative activities have a spontaneous aspect to them" (Gaut

[22] It is not for metaphysical reasons (that concern the alleged "reality" of things) that we treat these cases – as well as the relevant cases in the sections on the novelty and value conditions – as borderline cases. Rather, it is for pragmatic reasons – because the underlying issues cannot be settled here, and we take both sides of the disagreement seriously.

2018: 137). For that reason, Gaut adds spontaneity to his list of essential features of creativity. Similarly, Kronfeldner (2009, 2018) ascribes a pivotal role to spontaneity in the context of creative thought, conceiving of "spontaneity" as a "certain independence from the intentional control and the previously acquired knowledge of the person whose creativity is at issue" (Kronfeldner 2009: 579).

However, the view that creativity necessarily requires spontaneity is also problematic. Think, for example, about Piet Mondrian and his iconic grid paintings which consist of white, blue, red, and yellow rectangular spaces, separated by black horizontal and vertical lines. In the 1920s and 1930s, Mondrian painted numerous paintings in this style. According to the ignorance principle, a "creative person cannot have an exact plan of what she will do prior to being creative" (Gaut 2018: 135). Maybe this was true when Mondrian created his first-ever grid painting around 1920. However, it does not hold true when it comes to his later grid paintings. He knew exactly what he was doing, knew what he wanted to achieve and how to do so, and intentionally controlled the creative process. It is imaginable that, prior to the first brushstroke, he already knew exactly what the painting he was about to paint would look like. It is possible that this is true of all the grid-based paintings Mondrian ever painted apart from the very first one. In that case, if the ignorance principle and the spontaneity condition hold true, all of Mondrian's grid paintings, except for the first one he painted, would *not* be creative. But barely anyone would agree with this conclusion. Therefore, we would do well to refrain from treating spontaneity as a necessary condition of creativity.

(h) Surprise Condition

Margaret Boden (2004: 2ff.) holds that, besides novelty and value, surprise represents another necessary condition of creativity. However, this claim can be contested by pointing to Mondrian's work again. For it seems like most of Mondrian's contemporaries who were

familiar with his style were not surprised by seeing, say, his 107th grid painting. Yet, we would not say that this undermines its creative status.

It has been demonstrated that imagination, embodied improvisation, intrinsic motivation, spontaneity, and surprise are not necessary conditions for creativity. Besides that, it has been shown that novelty, value, and agency are not indisputable prerequisites for creativity. This represents a major obstacle to finding a neat definition of "creativity" that lists necessary (and jointly sufficient) conditions. In chapters 3 and 4, we will think about how to overcome this issue. But current creativity research does not only concentrate on how to define the concept of "creativity"; it also focusses on how to explain creativity, i.e., how to characterise what happens when we are creative. In the next section, we will attend to the most influential explanations or theories of creativity.

2.3 The Creative Process: What Happens When We Are Creative?

In the current scientific and philosophical literature, there are four major approaches to theorising about the nature and mechanisms of the creative process, that is, the process that brings about creative outcomes. These "Big Picture" approaches attempt to explain what happens when we are creative in general terms. They try to identify a mechanism or structure that underlies all creative processes. While some of the theories that will be presented conflict with each other, others are compatible. Even though all four explanatory approaches have some merit, it will be shown that all of them exhibit certain shortcomings. In the psychological literature, it is a contentious matter whether and to what extent there is empirical evidence in support of the theories that will be discussed in the following. To test their empirical validity, it is necessary to operationalise their key concepts. Yet, how to do so is often also contentious. So, in this section, we will not deal with questions about empirical testability; rather, we will point to *conceptual* problems that precede queries about scientific evidence.

It has been noted that practically all influential contemporary theories of creativity "take as their main explanandum specifically *human* creativity, with the dizzying accomplishments of great artists and scientists in some cases serving as their primary exemplars" (Shevlin 2021: 4). In brief, if one looks at the landscape of creative research, "creativity appears to be uniquely human" (Carruthers & Picciuto 2014: 199). The theoretical approaches that will be discussed in this section are no exception – they all conceive of creativity as a predominantly or exclusively human phenomenon. To put it differently: the notion of creativity that underlies all four approaches is committed to Paradigm 2 thinking. In addition, the following theories seem to rest on the assumption that novelty and value are necessary features of creativity. Some even explicitly embrace the two-part definition of creativity according to which novelty and value are not just necessary, but also sufficient for creativity.

(a) Sequentialist Approaches

A first set of theories shares the assumption that creativity is the outcome of a specific sequence of mental or cognitive steps. For that reason, such accounts may be referred to as "sequentialist". There are three major sub-sets of sequentialist theories.

(i) Classical Incubation-Based Models

Arguably the first outline of a sequentialist account of creativity was presented by the German scientist and philosopher Hermann von Helmholtz. In 1891, at a banquet in celebration of his seventieth birthday, he delivered a speech which touched on the process of generating new ideas. Based on introspective reflections, he asserted that "favourable ideas come unexpectedly and effortlessly" ["günstige Einfälle […] treten […] plötzlich ein, ohne Anstrengung" (Helmholtz 1896: 15)], that is, not while working on the problem, but in situations when one does not actively think about it. Alluding to Plato, he added that new ideas occur "like an inspiration" ["wie eine Inspiration" (Helmholtz 1896: 15)]. However, unlike Plato, Helmholtz

does not believe that the gods are the source of this sudden inspiration, but "extended efforts of preparatory work" ["längere vorausgehende Arbeit" (Helmholtz 1896: 15)].

In the early twentieth century, the French mathematician Henri Poincaré presented a similar introspection-based account of the creative process, more specifically, mathematical problem-solving. According to Poincaré, the solutions to mathematical problems are the outcome of "unconscious work" and "present themselves to the mind in a sort of sudden illumination" or, striking a more Platonic tone, in a moment of "sudden inspiration[…]" (1910: 328, 330, 329). Yet, like Helmholtz, Poincaré also emphasised that such spontaneous and unexpected illuminations are not God-given but require a "preliminary period of conscious work which always precedes all fruitful unconscious labor" (1910: 333) that eventually leads to the moment of inspiration. In contrast to Helmholtz, however, Poincaré noted that "when a sudden illumination seizes upon the mind of the mathematician, it usually happens that it does not deceive him, but it also sometimes happens[…] that it does not stand the test of verification" (1910: 332). Therefore, Poincaré holds that it is necessary to test the new idea that presents itself to the mind of the mathematician in a moment of sudden illumination according to established criteria.

In 1926, the British psychologist Graham Wallas systematised Helmholtz's and Poincaré's introspection-based reports and developed a sequentialist model according to which the creative process consists of four stages, namely: *Preparation*, *Incubation*, *Illumination* / *Inspiration*, and *Verification*. After consciously investigating a certain problem (*Preparation*), the mind of the creative individual keeps unconsciously working on the task (*Incubation*) which leads to the sudden occurrence of the solution (*Illumination* / *Inspiration*) which must then be tested (Verification) (Wallas 1926: 80).[23] Today, even though slight changes have been

[23] A similar model was developed by Hadamard (1945).

suggested[24], Wallas's four-stage model of creativity still enjoys great popularity. One reason for this seems to be that the model matches the popular stories we keep telling about iconic creative breakthroughs. Archimedes, entrusted with the task of checking the authenticity of the golden crown of King Hiero II, had his *Eureka* moment while sitting in the bathtub. Similarly, after researching the orbit of the moon and planets for an extended period of time, Newton had a sudden *aha* moment in which he grasped the basic idea of his theory of gravitation when he, as legend has it, saw an apple falling from a tree.

What Wallas calles "Incubation" is the centrepiece of the above-mentioned accounts. Accordingly, what he, Poincaré, and Helmholtz call "inspiration" is the outcome of unconscious processing of information, rather than a matter of divine intervention à la Plato. No supernatural force, but the human mind, understood as the locus of "incubation", is the source of creative insight. Despite this "humanised" conception of inspiration that is in line with Paradigm 2 thinking, the notion of unconscious incubation has been criticised for still being too enigmatic to be of explanatory value. As Guilford puts it in his seminal American Psychological Association keynote: incubation-based accounts "tell[…] us almost nothing about the mental operations that actually occur. […] The belief that the process of incubation is carried on in a region of the mind called the unconscious is of no help. It merely chases the problem out of sight and thereby the chaser feels excused from the necessity of continuing the chase further" (1950: 451). Guilford himself aims to continue the chase further. (In fact, Poincaré tries to cast at least some light on what happens during the Incubation phase; more on that later.)

(ii) Cognitive Sequential Models

In line with his criticism of classical incubation-based views on creativity, J. P. Guilford's own account strives to demystify "the nature of the processes that occur during the latent period of

[24] Jacob Getzels (1964), for example, suggested adding a stage which precedes the Preparation stage and revolves around formulating or finding the relevant problem.

incubation, as well as before it and after it" (1950: 451). To accomplish this, Guilford (1956) differentiates between three cognitive operations, three ways of thinking, which represent the sequences of a three-stage creative process. First, there is what he calls *divergent thinking* which involves thoughts "going off in various directions" (Guilford 1956: 274) and the generation of a multitude of ideas. The next stage consists of *convergent thinking* through which the ideas produced by divergent thinking are "channeled or controlled in the direction of [a useful] answer" (Guilford 1956: 274) or solution. Finally, the answer or solution brought about by *convergent thinking* needs to be scrutinised. Therefore, Guilford adds a third stage which relies on *evaluative thinking*.

A similar account was developed by Finke, Ward, and Smith (1992; 1999). It is usually referred to as the "creative cognition approach" or "Geneplore model" which describes "creative activities […] in terms of an initial generation of candidate ideas or solutions followed by extensive exploration of those ideas" (Ward et al. 1999: 191). In other words, according to the Geneplore model, creativity involves the generation and exploration of new ideas. (As "exploration" encompasses evaluation it comprises what Guilford calls *convergent thinking* and *evaluative thinking*.) Idea exploration comes after idea generation; yet, if it turns out during the exploratory stage that the ideas produced during the generative stage are of little help with regards to the relevant task, it might be necessary to generate new ideas. In this case, one might start to "alternate between generative and exploratory processes, refining the structures according to the demands or constraints of the particular task" (Ward et al. 1999: 191).

(iii) Darwinian Model

Other creativity researchers draw on evolutionary theory to get rid of the mysterious notion of "incubation". Their Darwinian models try to explain creativity in terms of the evolutionary process of natural selection. According to Darwin, natural selection involves two stages: first, the blind generation of new variation through mutation and, secondly, the selection and

retention of those variants that have a positive effect on survival and reproductive abilities. Analogously, Donald T. Campbell, the originator of the Darwinian approach, describes creativity as a "blind-variation-and-selective-retention process" (1960: 380). In this view, creativity involves two consecutive phases: an initial stage of blind variation in which numerous ideas are generated and a subsequent stage which selects the most useful of the ideas produced during the first stage. Or as Simonton, the most influential defender of the Darwinian model today, puts it: "The creator must generate many different novelties from which are selected those that satisfy some intellectual or aesthetic criteria" (1999: 21).

Seqentialist models are a common point of reference in the literature on creativity. They do not only fit in well with popular stories (or myths) about extraordinarily creative achievements, but also match the introspection-based accounts many of us have when we think about how we have found creative solutions in the past. And even if the picture the seqentialist models paint is not in line with one's experience, that does not necessarily refute them as most of the processes involved are happening unconsciously. However, it is questionable if the sequentialist models discussed are of universal validity. Does *every* creative activity really involve passing through the sequence of stages these theories posit? There is reason to doubt that.

All the seqentialist models delineated above share the assumption that creativity involves some element of "ideation", as it were, the generation of new ideas or solutions. The *Incubation* stage in the Inspiration-based theory, the *divergent thinking* phase in Guilford's account, the generative phase in the Geneplore model, and the blind variation stage in the Darwinian account all have in common that they generate new ideas or solutions. While the generation of new ideas is certainly an important feature of many creative processes, there seem to be instances of creativity which do not centre on ideation processes. Think, for example, of musical improvisation again. It is uncontroversial to say that musical improvisations often

exhibit creativity. However, improvisation is not the outcome of ideation. It would be odd to think that the minds of musicians generate ideas that allow them to pull off an improvisation. Moreover, improvisation takes place in the moment, so to speak, and appears, therefore, too spontaneous to be subjected to some kind of mental evaluation.

Furthermore, sometimes creativity results from applying some kind of technique. Think, for example, of Jackson Pollock's legendary drip technique. Pollock produced many paintings by applying this artistic technique. However, it does not seem that this involves the generation of novel ideas. Even if Pollock invented (or "ideated") his drip technique in line with the steps of one of the sequentialist theories, this does not mean that *applying* the technique involves the generation of new ideas. So, thinking about the production of, say, Pollock's 115th drip painting mainly in terms of ideation and idea generation seems like a mischaracterisation. Nonetheless, many people would still say that Pollock's 115th drip painting is an instance of creativity.

(b) Combinatorial Approach

(i) Classical Combinatorial Models

Sequentialist approaches are often conjoined with another approach which may be labelled "combinatorial". To shed some light on what was later dubbed "Incubation", Poincaré speculates that "during a period of apparent rest and unconscious work" the mind unconsciously forms "new combinations" (1910: 333) of distinct mental elements. He adds that "[a]mong chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart" (Poincaré 1910: 333). Darwinian theorists such as Campbell (1960) and Simonton paint a similar picture of the blind variation stage; as Simonton puts it: creativity is a "*combinatorial [...]* process" (1999: 81). Similarly, combinatorial processes also play an important role in the "Geneplore" model (Finke et al. 1992).

In brief, combinational accounts conceive of creativity as the combination of previously unrelated elements from different domains. Or as Sarnoff Mednick, one of the first

psychologists to adopt a combination-based approach to creativity research, puts it: "we may […] define the creative thinking process as the forming of associative elements into new combinations which either meet specified requirements or are in some way useful" (1962: 221).

Arguably the most famous combinatorial theory was developed by Arthur Koestler (1964; 2009). For Koestler, "the creative act consists in combining previously unrelated structures in such a way that you get more out of the emergent whole than you have put in" (2009: 251). He coins the term "bisociation" to refer to such a combinatorial act of "cross-fertilization" which represents, in his view, "the essence of creativity" (Koestler 2009: 252). Koestler tends to refer to the unconnected structures which are synthetized through *bisociation* as "matrices", representing "any ability, habit, or skill, any pattern of ordered behaviour" (1964: 38). So, "bisociation" stands for the combination of two matrices. To illustrate that, Koestler uses, amongst others, the examples of Gutenberg and Kepler: Gutenberg invented the printing press by combining two separate matrices, namely, the skill of printing with carved woodblocks and the skill of pressing wine; Kepler invented his laws of planetary motion by synthesising the matrices of physics and astronomy.[25]

Koestler maintains that "matrices" are "governed by a 'code' of fixed rules" (Koestler 1964: 38). For Koestler, "all coherent thinking and behaviour is subject to some specifiable code of rules to which its character of coherence is due—even though the code functions partly or entirely on unconscious levels of the mind, as it generally does" (1964: 42). However, even though the rules that structure matrices are fixed, matrices themselves allow for some flexibility. This is because, as Koestler argues, matrices can be adapted to the conditions of the environment. Therefore, as he points out, the exercise of a skill is not just governed by the fixed code of rules of the relevant matrix, but, at least to some extent, also by "a flexible strategy,

[25] As far as the *full* process of creative production is concerned, Koestler paints a picture which is similar to the Incubation-based accounts discussed above. In a sense, Koestler's combinatorial theory aims to explain what happens during the Incubation phase.

guided by environmental pointers" (Koestler 1964: 38). To illustrate that, he invokes the example of a spider that builds a web: the spider will construct the web according to "a fixed *code of rules* built into the spider's nervous system" (Koestler 1964: 38), whereas its choices of appropriate points of attachment in the environment is a matter of flexible strategy. Even though strategies allow for some flexibility of matrices, its code of rules sets a limit to their flexibility. So, the code of rules, rather than potential strategies, define matrices and equip them with their characteristic sense of coherence. For Koestler, bisociation combines the rules of two separate matrices: "The bisociative act means combining two different sets of rules" (Koestler 2009: 253).

(ii) Computational Combinatorial Models

Koestler's theory continues to have a significant impact on current creativity research. Gilles Fauconnier and Mark Turner, for example, draw on his work and reframe it in computational and cognitive science terms. Rather than speaking of "bisociation", the key notion of their model is what they call "conceptual integration" or "conceptual blending". They conceive of "creativity and novelty as consequences of conceptual integration" (Fauconnier & Turner 2003: 85; 2002: 382; also see Turner & Fauconnier 1999). For Fauconnier and Turner, conceptual blending or integration is a basic mental operation that "blends" different "mental spaces" so that a new structure emerges. Such "mental spaces" are defined as "very partial [mental] assemblies constructed as we think and talk for purposes of local understanding and action" (Fauconnier 2010: 351). Therefore, they are much smaller or "local" and more ephemeral than Koestlerian "matrices". Yet mental spaces are linked to more permanent mental structures: "A mental space is built up in part by recruiting structure from (possibly many) conceptual domains and from local context" (Fauconnier & Turner 1998: 157; also see Fauconnier 2010: 352). So, while mental spaces are generated in the working memory, they are nevertheless often connected to information stored in the long-term memory, that is (among other things) the

knowledge of what they call "conceptual domains". Fauconnier and Turner outline a complicated framework of how different mental spaces can be blended by combining their elements in various ways.

(iii) Neurocombinatorial Models

Apart from cautious speculations (Fauconnier 2010: 351), Fauconnier and Turner make no definite statements about how mental spaces, conceptual blending, and, thus, creative processes are linked to neurophysiological processes. However, other psychologists, neuroscientists, and philosophers are less cautious when it comes to postulating a neurophysiological basis of creative combinatory processes. Donald O. Hebb, for example, argues that "cell-assembly groups […] fire and subside, fire and subside, fire and subside, till the crucial combination occurs'' (Hebb 1980: 119). Moreover, even though this is sometimes overlooked, Koestler's account is, at its core, a neurocombinatorial one: the codes of rules that govern matrices refer, at the end of the day, to "structures in the nervous system" that "function on the trigger-release principle, so that a relatively simple signal-pattern releases complex, pre-set action-patterns" (1964: 631).

Drawing on Hebb, Koestler, and computational combinatorial models alike, Paul Thagard and Terrence C. Stewart develop a systematic neurocomputational approach according to which "mental representations are patterns of neural activity" (2011: 2). Accordingly, they conceive of creativity as "the combination of previously unconnected mental representations constituted by patterns of neural activity" (Thagard & Stewart 2011: 1). In brief, they hold that creativity results from combining existing neural patterns into new and valuable ones.

Although combinatorial theories are very popular in current creativity research, they are also problematic. One might object that "[t]here is no significant sense [...] in which Verdi's Requiem, or *Hamlet*, are combinations of previously existing elements or ideas" (Dartnall 1994:

537). However, we do not know the full mental or neural histories that led to the creation of these artistic achievements. These histories are not even accessible to the artists themselves since, according to practically all combinatorial theories, creative combinations are generated unconsciously. So, the defenders of the combinatorial approach might contend that the ideas that led to the creation of *Hamlet* might well have been the result of unconscious combinatorial processes.

Even if Verdi's *Requiem* and Shakespeare's *Hamlet* do not represent insurmountable challenges to the combinatorial approach, there are other counterexamples such as, again, the creation of Pollock's 115th drip painting. Many people would say that this artwork represents a creative achievement. To account for the creativity of Pollock's 115th drip painting, supporters of the combinatorial view must assume that, when Pollock applied his painterly technique to create this piece, he synthesized different elements from different frames of reference in his mind. Yet, this seems to mischaracterise what happened while Pollock created his 115th drip painting. Pollock's drip technique represents a pattern of ordered behaviour, something that can be "mastered". So, if one adopts a Koestlerian perspective, it makes more sense to characterise it as the application of a *matrix*, rather than the *fusion* of different matrices. Similarly, viewed through the lens of Fauconnier's and Turner's account, it seems implausible to assume that, every time Pollock applied his drip technique, the relevant mental structures were spontaneously generated in his working memory through conceptual blending. Again, it is important to differentiate between the *invention* of Pollock's drip technique and its *application*. So, even if combinatorial processes were involved in the former, this does not mean that they are essential for the latter. Consequently, combinatorial theories cannot account for the creativity that results from *applying* artistic techniques.

Besides that, there is another issue: combinatorial accounts paint an ahistorical picture of creative processes. According to them, creativity is the outcome of combining different mental or neural patterns – but they fall short of elaborating on how such creative outcomes

feed back into these kinds of patterns. Some creative achievements change the domains they are part of. Shakespeare's works, for example, changed the history of literature which means, based on the combinatorial models discussed above, that they changed the mental or neural patterns linked to the domain of literature. Yet, combinatorial theories fall short on explaining how these the mental or neural patterns change – and, therefore, fail to do justice to domain-changing creative achievements. Koestler briefly touches on this matter: even though he maintains that the rules of a matrix are fixed, he notes that "no rule is absolutely final" and that, under certain circumstances, rules "may be altered and combined into a more sophisticated game" (1964: 631). This seems to explain how matrices and the rules that structure them can change. Unfortunately, Koestler does not explore this further and does not discuss how the "alteration" of a rule and the formation of a "more sophisticated game", that is, of a more sophisticated matrix, relates to creativity.[26]

(c) Boden's Computational Approach

Margaret Boden's highly influential theory is arguably the most important point of reference in the current philosophy of creativity. It is also a "computational" theory as it rests on the assumption that "[c]omputational ideas can help us to understand how human creativity is possible" (Boden 2004: 17). Boden differentiates three distinct types of creativity which are the result of three different types of mental processes. The first type which she labels "combinational creativity" is, as the name suggests, akin to Koestler's notion of *bisociation* and Fauconnier's and Turner's notion of *conceptual blending*. Combinational creativity involves "making unfamiliar combinations of familiar ideas" (Boden 2004: 3; also cf. 2009a: 240). Yet, unlike combinatorial theories, Boden does not explain creativity solely with reference to

[26] In line with our criticism, Fauconnier and Turner (2003: 85) recognise that conceptual blending is not sufficient to fully explain creativity. Similarly, Thagard and Stewart (2011: 23) acknowledge the limitations of their neurocombinatorial approach and concede that it needs to be complemented by other factors and principles.

combinatorial processes but complements combinational creativity with two further types of creativity.

In order to specify the other two types of creativity, Boden introduces what she calls "conceptual spaces" which stand for structured sets of mental representations, for mental "maps inside our heads" (Boden 2004: 5). This may sound similar to the "mental spaces" in Fauconnier's and Turner's theory, but this is misleading as Boden's "conceptual spaces" are more comprehensive, more permanent, and less local than their "mental spaces". (In fact, Boden's "conceptual spaces" seem to be more similar to what Fauconnier and Turner call "conceptual domains".) As examples of conceptual spaces, Boden mentions, amongst other things, chess, physics, chemistry, story writing, Impressionism, or Baroque music. Accordingly, Boden emphasises that conceptual spaces "aren't originated by one individual mind" (2004: 4), but "normally picked up" (Boden 2004: 4), usually from one's own culture and peer group. Even though Boden's conceptual spaces vary from Koestlerian matrices insofar as they are mental rather than behavioral patterns, she concurs with Koestler in maintaining that conceptual spaces are structured by certain rules: it is "rules, conventions, or constraints that define [such thinking] styles" (Boden 2009a: 242). So, in short, conceptual spaces are rule-governed mental representations of specific domains or styles which are familiar to specific social groups.

On the basis of the notion of "conceptual space", Boden defines the two remaining types of creativity. In her view, conceptual spaces are not essentially static since, in contrast to Koestler, she contends that the rules that define a conceptual space "can be slightly altered or 'tweaked'" (Boden 2009a: 242). This happens in what Boden refers to as "exploratory creativity": "the existing stylistic rules or conventions [of a given conceptual space] are used to generate novel structures (ideas)" (2009a: 241) within that conceptual space. Less technically put: "exploratory creativity" happens when someone "comes up with a new idea within [a certain] thinking style" (Boden 2004: 4), i.e., discovers new structures or ideas within a pre-

existing conceptual space. According to Boden, the bulk of creative contributions to the arts and sciences represent this kind of creativity (2009a: 241). However, there are also acts of creativity that do not only explore, but radically transform conceptual spaces. Accordingly, she argues that there is a third kind of creativity which she calls "transformational creativity", and which consists in "changing the existing rules to create a new conceptual space" (Boden 2004: 58). So, in transformational creativity, "some deep dimension of the thinking style, or conceptual space, is altered—so that structures can now be generated which could not be generated before, and which are not all of a piece with the previous style" (Boden 2009a: 242). Boden maintains that transformational creativity is very rare and normally "valued more highly than the other two types" (Boden 2009a: 243).

Unlike the theories discussed above, Boden manages to account for the creativity of jazz improvisations and Pollock's 115th drip painting. According to her theory, jazz improvisation and Pollock's drip paintings represent distinct conceptual spaces. So, a *particular* jazz improvisation represents an exploration of the conceptual space of jazz improvisation; likewise, Pollock's 115th drip painting represents an exploration of the conceptual space of Pollock's painterly style. (Whether that is a good way of accounting for the creativity of musical improvisations and Pollock's 115th drip painting will be discussed later.) Furthermore, by introducing the notion of transformational creativity, Boden's model overcomes the ahistorical perspective of exclusively combinatorial approaches as it offers an account for how some creative acts can change conceptual spaces.

However, Boden's computational theory is also not flawless. Her model does not sufficiently account for the circumstance that new conceptual spaces can sometimes emerge which, even if they draw on elements of established conceptual spaces, are distinct from these. Accordingly, the creative emergence of such a new conceptual space is usually (instantly or retrospectively) perceived as the beginning of something new which defies existing categories. Using Boden's terminology, such creative achievements are, therefore, best characterised as

"developing a conceptual space *ab initio* rather than transforming an existing conceptual space" (Novitz 1999: 74). Take, for example, the invention of cinema. Even if the conceptual space of cinema draws on certain elements of older artistic conceptual spaces (e.g., specific narrative or directorial structures), it is recognised as an independent artform or conceptual space with its own rules, distinct from other conceptual spaces like the ones linked to theatre, the novel, or photography and their rules. Therefore, while it is apt to say that, for instance, Brecht's epic theatre transformed the conceptual space of theatre, it is a stretch to conceive of the emergence of cinema as a transformation of the conceptual space of theatre. So, the invention of cinema is better characterised as the formation of a new conceptual space.[27]

There is also a more general argument for introducing a kind of creativity that revolves specifically around the formation of new conceptual spaces. For, if we assume like Boden that a new conceptual space comes into being through the transformation of another conceptual space, this older conceptual space must have also come into being through the transformation of an even older conceptual space, and so on. To prevent an infinite regress, it is necessary to allow for the possibility of inventing new cognitive spaces that defy being classified as transformed versions of pre-existing cognitive spaces. (These are not the only problem concerning Boden's theory; a more substantial issue regarding her notion of "conceptual space" will be addressed in the next chapter.)

(d) Socio-Contextual Approach: Csíkszentmihályi's Systems Model

The above-mentioned theories presuppose that creativity is chiefly the outcome of mental or neurophysiological processes that take place inside the head of an individual. Incubation, divergent thinking, idea generation, blind variation, bisociation, conceptual blending, and the combination, exploration, and transformation of cognitive spaces are all processes that happen

[27] Novitz (1999: 73f.) presents a similar critique with reference to the emergence of the conceptual space of vaccinations or immunity through Edward Jenner's development of the smallpox vaccine.

inside the mind or brain of a single person. The underlying assumption that individual minds or brains are the locus of creativity has been criticised for neglecting the importance of social factors and the larger context within which creativity takes place.[28] Accordingly, some theories strive to overcome the individualistic perspective of the models discussed so far by pursuing a "socio-contextual" approach.[29] Socio-contextual models aim to do justice to the circumstance that "much of creativity emerges from complex social and organizational systems" (Sawyer 2012: 419).[30]

Arguably the most influential socio-contextual theory of creativity was developed by Mihály Csíkszentmihályi. His account is known as the Systems Model as it assumes that creativity is "a systemic rather than an individualist phenomenon" (Csíkszentmihályi 2013: 23). The systems model rests on the idea that creativity emerges from the interaction of a system which is constituted of three main components: first, there is what Csíkszentmihályi calls the "domain" which stands for specialised areas of activity, things like geometry, music, painting, business, engineering. Csíkszentmihályi asserts that "[e]ach domain is made up of its own symbolic elements, its own rules, and generally has its own system of notification" (2013: 37). In other words, the domain exhibits standards of correctness, it is governed by rules. Secondly, a system entails a "field" which represents the "gatekeepers to the domain" (Csíkszentmihályi 2013: 28). The experts that are part of the field decide which ideas or products are included in

[28] Similar criticisms against overly individualistic accounts of creativity and invention have already been voiced in the 1930s by Lev Vygotsky (2004: 10f.).

[29] Even if it does not represent a wholly socio-contextual approach, Amabile's (2013) Componential Theory holds that creativity depends on a confluence of several psychological components and a social component, i.e., an adequate social environment. One of the psychological components she invokes is intrinsic task motivation. As the Componential Theory rests on the idea that the components it lists are necessary and sufficient for creativity (Amabile 1983: 357), it presupposes that intrinsic motivation is necessary for creativity. However, in section 2.2., we have seen that this assumption, a pillar the Componential Theory rests upon, is unfounded.

[30] The "Creative Audience Account" which has been proposed by Carroll (2014) also criticises traditional creativity research for mainly concentrating on the mental activities of creative people. In Carroll's view, audiences co-create artworks. However, the explanatory scope of the Creative Audience Approach is limited as it solely deals with *artistic* creativity. For, as far as creativity in everyday situations is concerned, there is normally no audience one can or needs to creatively cooperate with. Furthermore, Carroll holds that artistic creativity requires imagination both on the side of the creator and the side of the audience (2014: 63). However, as shown in section 2.2, there are cases of artistic creativity that do not rely on imagination.

the domain and which are not. In the art world, curators, critics, collectors, and museums are important parts of the field; in academic domains, the field comprises journals, universities, research committees, and research councils, among other things. The third, and final, component of the system is the individual "person" that works in a certain domain. For Csíkszentmihályi, creativity is a social construct that is "jointly constituted by the interaction among domain, field, and person" (2013: 29). According to his systems model, creativity happens when a person, by drawing on the rules of a given domain, produces some novelty that is appreciated by the relevant field and considered worthy of inclusion into the respective domain.

Even though Csíkszentmihályi acknowledges that "some of the most creative breakthroughs occur when an idea that works well in one domain gets grafted to another" (2013: 88), he does not maintain that this represents the only way of being creative. So, the criticisms levelled against the combinatorial approach do not apply here. Furthermore, Csíkszentmihályi notes that "[o]ccasionally creativity involves the establishment of a new domain" (2013: 28); yet, he does not adopt Boden's position that this can only happen by transforming another domain. Accordingly, Csíkszentmihályi states that "there was a time when domains did not exist. The first astronomers, the first chemists, the first composers were not changing a domain but actually bringing one into being" (2013: 291).

Despite its virtues, the systems model also exhibits certain shortcomings. Csíkszentmihályi defines a "creative person" as "someone whose thoughts or actions change a domain, or establish a new domain" (Csíkszentmihályi 2013: 8). Of course, "changing a domain" can mean many things, but the changes Csíkszentmihályi has in mind are major paradigm-shifting transformations "that change[…] some aspect of the culture" (Csíkszentmihályi 2013: 27). (In a sense, the changes he talks about seem to be similar to the changes brought about by transformational creativity in Boden's account.) However, most people neither change, nor establish a new domain in their lifetimes. So, while the conception

of creativity that underlies Csíkszentmihályi's theory might apply to the likes of Galileo, Mozart, or Picasso, it does not account for less epochal instances of creativity. Again, performing a standard jazz improvisation (i.e., one that does not radically transform the domain of jazz) would not qualify as "creative". Consequently, the systems model rests on an overly narrow and highly elitist conception of "creativity". One needs to be a genius, if you will, to be creative. To be fair, Csíkszentmihályi mentions that his account mainly tries to come to grips with what psychologists commonly refer to as "Big-C" creativity, that is, the creativity of extraordinarily creative persons (2013: 8, 27). But this only confirms that the systems model fails to do justice to the diverse manifestations of creativity.

On a related note, the systems model – which has been compared to the Institutional Theory of Art[31] (Gaut 2010: 1037) – overemphasises the role of what Csíkszentmihályi labels "field", of the gatekeepers, experts, and established institutions of a domain. At the end of the day, it is the field that decides whether an idea or object gets incorporated into the domain. Thus, the field determines what qualifies as "creative" and what does not. However, the views and valuations of the field are not always universally shared within a domain. Many domains comprise a vast variety of, partly, opposing views, opinions, and assessments, with some being more widely shared than others. However, the systems model declares all those voices that do not coincide with the authoritative evaluations of the field to be irrelevant when it comes to determining what is creative.[32] This underscores the elitism implicit in the systems model. It fails to do justice to the diversity of views and evaluations within a domain. (Besides these problems, the next chapter will expose an even more fundamental issue concerning Csíkszentmihályi's model.)

[31] Simply put, the Institutional Theory of Art defines "art" as what museums, galleries, and other institutions of the artworld accept as "art".

[32] Furthermore, it has been argued that, due to the impact of modern information and communication technologies, the influence of the traditional gatekeepers Csíkszentmihályi places at the heart of the "field" has decreased over the last few decades (Gangadharbatla 2010).

Of course, there are more models of creativity than the ones discussed above. However, these are the most influential ones in the philosophical literature and beyond. Often, "nomothetic (or universalist) emphasis" (Kozbelt 2011: 477) is put on these theories by their proponents or the creativity research community more generally. However, it has been shown that all these theories fail to account for certain examples of creativity. In other words, they neglect the diversity of the phenomenon of creativity. Wittgenstein famously said that a "main cause of philosophical disease" is "a one-sided diet: one nourishes one's thinking with only one kind of example" (PI: §593). Even if the models scrutinised above might not be preoccupied with just one kind of example, they are still biased towards certain kinds of examples, while they disregard others. This calls for a new way of thinking about "creativity".

3. Conceptualising Creativity: A New Perspective

In the previous chapter, it has been demonstrated that widespread attempts to define and explain creativity exhibit certain shortcomings. In this chapter, we will set the scene for a novel approach to examining creativity.

3.1 Creativity Beyond Essentialisms and Sharp Boundaries

Many philosophers from Plato to Frege and beyond hold the view that concepts can and have to be defined by invoking a finite number of essential features (or conditions) which are shared by all and only the entities to which the concept applies. This view on "concepthood", as it were, underlies the current definitional debates about creativity: contemporary philosophy of creativity is driven by the attempt to find essential features of "creativity". However, an essentialist definition of "creativity" appears to be beyond our grasp, as, in section 2.2, we could not detect a single feature that is indisputably shared by all cases of creativity.

We have seen that novelty, value, and agency are contentious contenders for necessary conditions for creativity. So, although they come closest to representing essential features of "creativity", any definition of "creativity" that contains all or some of these conditions is, at the end of the day, as controversial as they are. But let us brush our doubts aside for a moment and treat novelty, value, and agency as if they were uncontentiously necessary conditions of "creativity". Do these three features capture the essence of "creativity"? An essentialist or "analytical" definition of concept x does not only require the set of conditions it entails to be shared by all the things falling under concept x – it also demands that this set of conditions is *only* exhibited by things belonging to concept x. In other words, an analytical definition needs to invoke conditions that are both necessary *and* jointly sufficient. As Frege puts it: "A definition of a concept […] must be complete" (1960: 159). (In the following, I will use the labels "analytical", "essentialist", and "Fregean definition" interchangeably.)

However, it can be shown that, even if we assume that novelty, value, and agency are necessary for creativity, they are *not* jointly sufficient for, or exclusive to, it. Take, for example, the phenomenon of daydreaming. Some daydreams are new insofar as neither the daydreamer, nor anyone else, ever has had these exact daydreams before. (If one holds on to the novelty condition, a weak form of newness suffices here.) Daydreams can be pleasant – and, thus, valuable – to the dreamer. Moreover, if one accepts a weak notion of agency (which is, as shown in section 2.2, necessary if one wants to adhere to the agency condition), the act of daydreaming involves some degree of agency as the daydreamer has at least some power over the daydream and its contents. So, one can conclude that such a daydream involves novelty, value, and agency. However, we would not say that all such daydreams are creative. So, novelty, value, and agency are not only contentious necessary conditions, but also not jointly sufficient, for "creativity".

According to Frege's influential conception of "concepthood" that implicitly underlies the vast majority of philosophical attempts to define "creativity" today, "a concept that is not sharply defined is wrongly termed a concept" (1960: 159). Since a definition that lists necessary and jointly sufficient conditions and, in doing so, sharply demarcates its boundaries is not available – do we have to conclude that "creativity" is wrongly termed a concept and that we cannot make any sense of it? No, we don't – and Wittgenstein helps us to realise this.[33]

After Wittgenstein endorsed similar ideas in his first book, the *Tractatus Logico-Philosophicus*, he rejected the essentialist picture of language and concepthood in his later writings, especially, in his *Philosophical Investigations*. First, he reminds us that our language hosts many words and concepts that resist being defined in an analytical way, that is, in terms of necessary and jointly sufficient conditions. He famously points to the concept of "game" and emphasises the diversity of the things this noun applies to: there are board games, card games,

[33] Another influential critique of the take on concepthood underlying Frege's remarks can be found in Quine (1951).

ball games, martial games, and many others; some games are competitive, others are not; some are entertaining, others are not: "if you look at them you will not see something that is common to *all*" (PI: §66). At the same time, there are activities that are competitive (like management) or entertaining (like watching a film), without being games. Therefore, it appears to be hopeless to find something all and only games share.[34] Besides "game" (PI: §66f.), Wittgenstein maintains that "language" (PI: 65ff., 92, 108), "number" (PI: §§67f.), and "reading" (PI: §§156-78) are further examples of concepts that defy being analysed in terms of necessary and jointly sufficient conditions.

There are two common misconceptions about Wittgenstein's attack on the Fregean take on concepthood and essentialism more generally. First, it is easy to read Wittgenstein in a way that suggests that *all* concepts defy being analysed in terms of necessary and jointly sufficient conditions. Regardless of whether this is a misreading or not, this claim does not withstand scrutiny on systematic grounds. Consider, for example, the concepts of "octagon"; having exactly eight angles is necessary and sufficient for a geometrical object to be an "octagon". Or take the concept of "tricolor" (Ahmed 2010: 47); having three bands of different colours is necessary and sufficient for a flag to be a "tricolor". Consequently, as these examples illustrate, there are concepts that can be analysed in terms of necessary and jointly sufficient conditions.[35]

Secondly, it is a misunderstanding to claim that there *exist* no necessary and jointly sufficient conditions for concepts like "game", "language", "number", or "reading", that is, for concepts we fail to analyse in terms of necessary and jointly sufficient conditions. Statements about the inexistence of necessary and jointly sufficient conditions are metaphysical claims that cut across the avowed anti-metaphysical (PI: §116) and anti-dogmatic (PI: §131) thrust of Wittgenstein's later philosophy. On a systematic basis, it is also advisable to adopt a more

[34] One might argue that there is something all "games" have in common, namely that, they all involve some sort of (physical or mental) activity. Yet, this also pertains to phenomena which are not "games" such as housebuilding, war, or writing a thesis. Thus, the suggested criterion is, at most, a necessary, but no sufficient, condition.
[35] For a different critique of the view that Wittgenstein maintains that *all* concepts defy being analysed in terms of necessary and jointly sufficient conditions, see Sluga (2011: 80).

agnostic position that avoids metaphysical assumptions about indefinability (in virtue of necessary and sufficient conditions). For, maybe, a superintelligent being might be able to compile millions of predicates that, in conjunction, successfully capture the essence of "games". However, this would not change anything; for we can and actually do employ and understand words like "game" in everyday life without consulting such a catalogue of necessary and sufficient conditions. In practice, we do not need analytical definitions to be able to use words like "game" correctly and grasp their meaning. Thus, by rejecting the Fregean account of concepthood, one is "not committed to the nonexistence of a common essence but rather to the irrelevance of such a commonality to the use of an expression and therefore to its meaning" (Ben-Yami 2017: 411). For that reason, we should be agnostic about whether concepts like "game" are really undefinable in terms of a list of necessary and sufficient conditions, a list of features that, if it exists, is probably beyond the cognitive grasp of human beings. What is important is that such a list is not available to us human beings and, thus, not necessary for us to use and understand these terms.

Even though we cannot identify a set of essential features that are common to all, and only, the things we call "games", Wittgenstein writes that it is still possible to discern a "complicated network of similarities overlapping and criss-crossing" (PI: §66). To characterise the manifold similarities and affinities that obtain between the things we apply the concept of "games" to, Wittgenstein introduces the notion of "family resemblance" (PI: §67), alluding to the "various resemblances between members of a family" (PI: §67). Even if we cannot specify essential features shared by all, and only, the things a family-resemblance concept applies to, there are overlapping resemblances between these things.

Despite the fact that there is no practically feasible and comprehensible analytical definition of "creativity" available to us, we usually have no problem understanding and using this concept in everyday contexts. Should we, therefore, conclude that "creativity" is best

understood as a family-resemblance concept?[36] A problem with Wittgenstein's notion of "family resemblance" is that he says very little about the positive features of this kind of concept and that his use of this metaphor sometimes lacks coherence (Forster 2010: 69). This leaves plenty of space for disagreement as to what being a family-resemblance concept actually entails for Wittgenstein and how to characterise the "similarity relations" that obtain between the things that are covered by this kind of concept.

Unless one presupposes a restrictive conception of "similarity", it seems uncontroversial to say that there are similarities between the things we call "creative". Moreover, most people would probably agree that, among the things we call "creative", some things exhibit greater similarities than others (just as some games are more similar than others). For example: from a certain angle, the creative achievements of Copernicus or a Darwin resemble each other, despite certain differences, insofar as they provided "a fertile new point of view" (Wittgenstein 1980a: 43e) that gave their respective scientific domains a radically new impetus – which is something that cannot be said of all, in fact only of very few, of the things we call "creative". On some interpretations of "family-resemblance", this might suffice to classify "creativity" as a family-resemblance concept; on other interpretations, it might not.

In the light of such disagreements, it is better to abstain from using the label "family resemblance" when characterising the concept of "creativity".[37] (At a later point, we will characterise some of the similarities that obtain between various instances of creativity in greater detail – but we will do so independently of the question of whether "creativity" is best described as a family-resemblance concept.) For our purposes, it is enough to acknowledge that the essentialist or Fregean approach of detecting necessary and sufficient features turns out to be a dead-end when dealing with the concept of "creativity". Interestingly, Margaret Boden,

[36] Mogensen (2018: 57) and Dresler (2008: 15f.), for example, classify "creativity" as a family-resemblance concept, even though they remain short on an in-depth analysis that justifies this characterisation.
[37] Similarly, Erden (2010) and Grève and Mácha (2016: 4) express that they are sceptical of the attempt to define "creativity" in an essentialist manner but desist from categorising "creativity" as a family-resemblance concept.

despite her oft-cited analytical definition of creativity in terms of novelty, value, and surprise, more recently remarked that she "repudiate[s] essentialist approaches" and that "[h]ard- and-fast definitions generating dichotomous distinctions aren't available […] for creativity" (2014: 233). Nevertheless, despite her avowed commitment to anti-essentialism, Boden still sticks to her influential analytic definition of creativity (2014: 227).

From a Fregean perspective, the concept of "creativity" is not only peculiar for defying (practically viable) analytic definitions. For Frege not only insists that concepts need to be analytically definable, but also that the definition of a concept "must unambiguously determine, as regards any object, whether or not it falls under the concept" (1960: 159). In brief, for Frege, concepts require sharp boundaries. A concept has sharp boundaries if it is always clear whether this concept applies or not. So, a concept lacks sharp boundaries if there are borderline cases. Yet, as shown in section 2.2, there are not only contentious conditions for, but also borderline cases of "creativity" – cases with regards to which it is not clear whether they fall under the concept of "creativity" (e.g., using the stream-of-consciousness technique in 1985 or Tesla's wayward views on spacetime). Logicians tend to refer to concepts that lack sharp boundaries as "vague": "There is wide agreement that a term is vague to the extent that it has borderline cases" (Sorensen 2018).

Many concepts that defy being defined in terms of necessary and jointly sufficient conditions (in a practically viable and comprehensible way) are also vague. The concept of "game" helps us to illustrate that once again. Think, for example, of "a child [that] throws his ball at the wall and catches it again" (PI: §66). Is this child playing a "game"? Some might say yes, some might say no, others might be uncertain. However, not all concepts that (practically) defy being analytically defined are also vague. As noted, the concept of 'number' is an example of a concept we (practically) fail to define analytically (PI: §67); yet, it is not vague as "there are no 'borderline numbers' in that sense in which there might be borderline blue objects" (Ahmed 2010: 47). At the same time, a concept can be vague, but still be analytically definable.

As mentioned, "tricolor" is a concept that can be analysed in terms of necessary and jointly sufficient conditions. However, the concepts of popular tricolour stripe colours such as "red" or "blue" admit of borderline cases so that the concept of "tricolor" as a whole is vague: "it is easy to imagine a borderline tricolor" (Ahmed 2010: 47). Thus, being vague and being (practically) indefinable in terms of necessary and jointly sufficient conditions are not mutually dependent.

The Fregean account of concepthood according to which concepts require sharp boundaries implies that vague concepts are not concepts at all. To justify that claim, Frege compares a concept without sharp boundaries with "an area that had not a sharp boundary-line all round, but in places just vaguely faded away into the background" and contends that "[t]his would not really be an area at all" (1960: 159). Wittgenstein refers to this passage and counters it by posing the following question: "is it senseless to say: 'Stand roughly there'?" (PI: §71). Wittgenstein denies this; to make that point, he provides the following scenario: "Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand—as if I were indicating a particular spot" (PI: §71). The quoted passage, as I read it, highlights that, in many practical contexts, it is not possible to maintain the ideal of sharpness and clarity that underlies Frege's remarks. The world we live in often exhibits a significant amount of complexity and multifacetedness and can be subject to change. That often makes it difficult, if not practically impossible, to draw sharp boundaries. Consequently, in practice, it is not always evident whether a certain concept applies or not. To avoid potential misunderstandings: we do not claim that *all* concepts are vague. (As noted, there is, for example, reason to assume that the concept of "number" is not a vague concept.) We only say that *some* of our concepts are vague, namely those which occasionally give rise to disagreement or uncertainty about their application.[38]

[38] As with respect to the analytical definability of concepts, we do not make any claims about whether there are metaphysical facts about the boundaries of vague concepts or not. We only say that, in practical contexts, some

We have seen that there is sometimes substantive disagreement or uncertainty about whether something is "creative" or not. Thus, "creativity" is a vague concept. An adequate account of the "nature" of "creativity" must come to grips with the concept's vagueness and the unavailability of an analytical definition. In other words: in order to make sense of "creativity", we should acknowledge that there are borderline cases of "creativity" and must not fall back on attempting to find necessary and jointly conditions for the application of this concept. A satisfying account of "creativity" needs to cope with these two requirements.

However, not just the most prominent attempts to *define*, but also the most influential attempts to *explain* creativity exhibit serious limitations. As we have seen in section 2.3, they keep to an unbalanced diet of examples. So, not just the most prominent definitions of "creativity" tend to neglect the diverse forms and manifestations of creativity – the same is true of the most influential explanations or theories of creativity. In order to avoid the confusions and deficiencies which have been detected in the most influential theories of creativity, it is helpful to compile a list of requirements or expectations which an appropriate account of creativity should satisfy.

3.2 Requirements for a Framework of Creativity

In contrast to the definitions and explanations discussed so far, an adequate account of creativity should capture the diversity and multifacetedness of "creativity". In the previous section, we have already stated two requirements: a satisfactory account should, first, do justice to the circumstance that there is no analytic definition in terms of necessary and jointly sufficient conditions available to us and, secondly, acknowledge the existence of borderline cases of

concepts are vague. This position is not incompatible with metaphysical views about vagueness according to which vague concepts have sharp metaphysical boundaries. Epistemicism about vagueness, as defended by Williamson (1994), for example, holds that such sharp boundaries exist, but we do not know them. Yet, this view does not reject the existence of vagueness: "[t]he epistemic view does not deny that there is vagueness", but only provides "a theory of what vagueness is" (Williamson 1997: 921).

"creativity" in the face of the significant disagreement that occasionally arises when assessing whether something is "creative" or not. This leads us to the first two requirements:

(1) *Overcoming Essentialisms and Sharp Boundaries: A satisfactory account of creativity …*

(1a) *does not depend on positing necessary and sufficient conditions of "creativity".*

(1b) *acknowledges the existence of borderline cases of "creativity".*

Based on our discussion of the most influential theories of "creativity" in section 2.3, we can state further requirements. To get a grip on the phenomenon of creativity, most of these theories introduce certain *frames of reference* such as "matrices", "conceptual spaces", "domains", "patterns of neural activity", etc. These frames of reference are the foundations of the respective theories. There is a widespread sense that, as Heinrich Wölfflin put it, "not everything is possible in every period" (quoted from Gombrich 1962: 4). For the people of the Middle Ages, it seemed "impossible" that a black shape on a white canvas can be art. Yet, in the early twentieth century, when Kazimir Malevich painted his *Black Square*, this did no longer represent an impossibility. Similar things can be said about the sciences: centuries ago, it seemed "impossible" to think that the earth is not the centre of the universe or that humans evolved from primates. What we call the "frame of reference" defines what is possible at a certain point in time.

However, some instances of creativity change what is considered possible and impossible – by changing the frame of reference linked to the relevant domain. As shown above, combinatorial approaches do not elaborate on how creative ideas, actions, or artifacts sometimes transform a given frame of reference; thus, they are limited to an ahistorical, synchronic perspective on creativity. Sequentialist models, even though they explain creativity in terms of a diachronic sequence of mental or cognitive steps, adopt a similarly synchronic take on creativity by focusing on how new ideas are being generated, rather than on how new ideas relate to the wider structures or networks which serve as the background against which

new ideas are perceived as new. Thus, one should accommodate the circumstance that some creative achievements transform the wider frames of reference they are part of. Besides that, we have criticised the view underlying Boden's model that a new frame of reference (that is, in her framework, a new conceptual space) can only emerge through the transformation of a pre-existing frame of reference (conceptual space). For, as demonstrated, some new frames of reference (even though they might draw on certain elements of pre-existing frames of reference) defy being characterised as transformed versions of pre-existing frames of reference. At the same time, the framework should not paint an elitist picture of creativity according to which one is *only* creative if one transforms an existing frame of reference in a significant sense or invents a new one. So, one does not necessarily need to something radically new to do something "creative". Creativity can, for example, sometimes result from applying an established technique. These points suggest the following set of requirements:

(2) *Framing Creativity: A satisfactory account of creativity …*

(2a) *conceives of creativity not only synchronically, but also diachronically.*

(2b) *makes sense of the circumstance that some instances of creativity change existing frames of reference.*

(2c) *accounts for the idea that new frames of reference can emerge which, even though they might draw on elements of pre-existing frames of reference, are distinct from these and are, therefore, (instantly or retrospectively) perceived as the beginning of something new.*

(2d) *does justice to the circumstance that there are other ways of being creative apart from significantly transforming existing frames of reference or inventing new ones. In doing so, it manages to cope with the circumstance that creativity sometimes results from applying established techniques.*

These requirements ensure that a satisfactory framework does justice to the multifacetedness of creative processes, the circumstance that one can be creative in different

ways. In conjunction, they guarantee that creativity is not reduced to a single operation such as the combination of elements from different frames of reference as suggested by combinatorial theories. Furthermore, requirement (2d), by registering that creativity is sometimes the outcome of applying established techniques, ensures that "creativity" is not reduced to idea generation which helps to overcome the one-sided perspective underlying most sequentialist models.

Moreover, the socio-contextual approach is right in criticising overly individualistic, subject-centred conceptions of creativity. For normally, individuals are not creative in a social vacuum. Accordingly, collaboration is often essential for creative achievements: think, for example, of artist collectives like *les Fauves, The Blue Rider, Oulipo,* or the Bauhaus. The history of science, technology, and engineering also features many examples of collaboratively creative efforts. The creation of the digital computer is one of them. This becomes particularly evident when one looks at the development of the first stored-program electronic digital computer in history, the *Small Scale Experimental Machine* (aka the *Manchester Baby*). Freddie Williams, one of its creators, emphasises that collaboration was essential for this milestone in computer engineering: "Tom Kilburn and I knew nothing about computers, but a lot about circuits. Professor Newman and Mr. A.M. Turing […] knew a lot about computers and substantially nothing about electronics" (1975: 328).

Even creative achievements that are associated with only one person heavily rely on collaboration. We revere, for example, Einstein as the originator of the theory of relativity – and there is, of course, nothing wrong about this as his exceptional creativity and intelligence was a decisive factor of its development. However, even Einstein relied heavily on collaboration: he had friends and colleagues like Marcel Grossmann and Tullio Levi-Civita who helped him with the maths. Moreover, he availed himself of the work of Hendrik Antoon Lorentz (*Lorentz transformations*), Hermann Minkowski (*Minkowski space* for general relativity), and many others. Without their work, Einstein would not have been able to develop *Relativitätstheorie*. Einstein himself even explicitly expressed that his work "represents an

intellectual development to which many contributed in crucial ways" ["es sich hier um eine gedankliche Entwicklung handelt, an der Viele ganz wesentlich betieligt waren" (quoted from Seelig 1986: VIII)]. Lorentz, Minkowski, and all the others who paved the way for Einstein, in turn, also drew on the work done by other humans which, again, drew on the work of others, and so on. Therefore, as Csíkszentmihályi puts it, to say that "Einstein discovered relativity is a convenient simplification. […] To say that the theory of relativity was created by Einstein is like saying that it is the spark that is responsible for the fire. The spark is necessary, but without air and tinder there would be no flame" (2013: 7). This illustrates that creative achievements are normally not the creation of a single mind, but the result of socially embedded, scaffolded, and distributed processes.[39]

Besides that, both the subject-centred and socio-contextual models scrutinised in 2.3 tend to abstract creative processes from the environmental context they are embedded in. Yet, to frame it in allusion to Wölfflin's dictum, it seems that it is not only the case that not everything is possible in every period, but also that not everything is possible in every environment. This becomes evident when one considers that some of the differences between Western medieval and eighteenth-century musical works are due to the different surroundings, that is, the disparate environments, in which they were commonly performed (Byrne 2012: 16-21). Wheeler describes it as follows: the former "was standardly performed in huge stone-walled gothic cathedrals with long reverberation times [which] explains why [Western medieval] music evolved to feature modal structures characterised by very long notes"; eighteenth-century compositions like those of Mozart, on the other hand, were usually performed "in grand, but not gigantic, rooms populated by plenty of people in extravagant dress", that is, in a "sonic environment [that] deadened the sound and allowed elaborate details

[39] Here, we only state that our framework should account for the circumstance that creative processes are *often* social and distributed. We do not claim that they are necessarily socially distributed. Can, for example, a human being that has never been exposed to any other humans be creative? My inclination is to say yes. Yet, there is no space here to examine this question in detail. Therefore, we restrict ourselves to the weaker and less contentious position that creative processes are, at least, very often socially distributed.

in the music to be heard" (2018: 241). Referring solely to mental representations or neural constellations to explain these differences seems inadequate. So, frames of reference such as "mental spaces", "matrices", "conceptual spaces", "patterns of neural activity", etc. are too mentalistic or neuro-centric to capture the environmental embeddedness of creative efforts and achievements. While the frames of reference invoked by socio-contextual models such as Csíkszentmihályi's "domains" might be able to cope with social factors, they still have an insufficient grasp of the impact environmental factors have on creative efforts. This brings us the final set of requirements:

(3) *Social and Environmental Factors: A satisfactory account of creativity invokes a frame of reference that …*

(3a) *captures the socially embedded, scaffolded, and distributed character of many creative processes.*

(3b) *does not disconnect creative processes from the environmental context they are embedded in.*

The account of creativity that will be developed in the following aims to satisfy requirements (1) to (3).[40]

[40] Subject-centred theories of creativity tend to not only adopt an overly individualistic, but also a disembodied perspective on creativity as recently criticised by Wheeler (2018: 236ff.). Most sequentialist, combinatorial, and computational models focus on what is happening inside the head of a creative person; if there is reference to bodily factors, it is normally limited to neural structures. Even socio-contextual models are inclined to neglect the impact non-neural bodily features have in the context of many creative processes. However, even if one accepts that (non-neural) bodily factors shape creative processes, it requires a lot of theoretical effort to specify how exactly they do so, especially if one's theory rests on a mental frame of reference. Due to space limitations, these efforts cannot be pursued here. However, the framework that will be developed does not limit us to a disembodied perspective on creative processes; yet, at the same time, it bypasses thorny philosophical questions about how the mind relates to the body. Wheeler who is mainly interested in investigating the creative *mind* also draws on 4E approaches to cognition. Accordingly, with reference to the extended mind hypothesis (Clark & Chalmers 1998), Wheeler (2018: 243-247) not only argues that the creative mind is embodied and embedded, but also extended. Even though I think that the extended mind hypothesis has its virtues, I do not invoke a "mental externalism" requirement as this seems to presuppose that one conceives of creativity primarily as a mental phenomenon, even if cognition is not limited to the boundaries of the skull. Accordingly, a mental externalism requirement seems to presuppose a frame of reference that consists of mental elements. Yet, it has not yet been discussed if such a frame of reference is adequate with respect to our purposes.

3.3 The Rules of Creativity: Wittgenstein and Rule-Following

It is common to conceive of creativity in terms of rules. The literature on creativity is pervaded by statements like the following: "There are norms for how human beings act, of course. But creativity in the genuine sense is the ability to change those norms in some important human domain" (Kelly 2019: 73). When speaking about creative achievements – or, more precisely, highly influential creative achievements –, for many people, it is natural to say things like "the Cubists changed the rules of art" or "Arnold Schönberg broke with the rules of classical music". Three of the most influential theories of creativity that have been discussed in section 2.3 also analyse the creative process with explicit reference to rules, namely Boden's computational model, Koestler's combinatorial account, and Csíkszentmihályi's systems model. All three accounts give different explanations of what happens when one changes or "plays with" the rules of music, painting, poetry, chemistry, or any other creative domain. In so doing, they presuppose different foundations for the rules that are changed or played with in the creative process.

As we have seen, Boden's model explains creativity in terms of "conceptual spaces", that is, mental representations, or "mental map[s]" (Boden 2004: 5), of specific domains or styles. As outlined above, Boden holds that creativity is the result of a transformation (2004: 5f.; 2009a: 242f.), exploration (2004: 4f.; 2009a: 241f.), or unfamiliar combination (2004: 3; 2009a: 240) of such conceptual spaces that "exist in the mind" (2009b: 235). Therefore, on Boden's account, Schönberg, for example, by inventing the twelve-tone technique, changed the rules of Western music by changing the "conceptual space" of (classical) music. Since Boden maintains that conceptual spaces are governed by rules (2009a: 242), one can also frame the matter as follows: Schönberg changed the rules of Western music by changing the (mental) rules of the "conceptual space" of (classical) music.

Among the theories scrutinised in section 2.3, there are some that, even though they do not explicitly talk about "rules", posit mental foundations for the processes they deem essential

to creativity. Fauconnier and Turner, for example, invoke "mental spaces". As noted above, such "mental spaces" (which are described as rather local and evanescent) are often connected to broader "conceptual domains" which seem to be roughly equivalent to Boden's "conceptual spaces". In addition, sequentialist theories assume that creativity is the outcome of a process that involves the generation as well as the exploration, channelling, or selection of ideas or, more technically speaking, of mental representations. So, if one wanted to reframe these models in terms of rules, the rules would also be connected to mental representations and processes.

Like Boden, Koestler also explicitly speaks of rules. According to his theory, creativity results from an act of "bisociation" which combines the rules from different matrices (Koestler 2009: 253). However, Koestler's language is less "mentalistic" than Boden's. He characterises "matrices" in dispositional or behavioural terms, describing them as habits, abilities, skills, or routines (Koestler 1964: 38, 638f.). What is more, for Koestler, matrices and the codes of rules that structure them have a neurophysiological basis. He writes that the rules of matrices "refer to concrete processes or patterns of organization in the central nervous system" (Koestler 1964: 638). The rules of matrices such as "non-Euclidian geometry or quantum mechanics are physiologically represented[…] in the nervous system" (Koestler 1964: 638). So, according to Koestler's picture, Kepler, for example, invented his laws of planetary motion and thus changed the rules of science by combining the neurophysiological patterns associated with the matrices of physics and astronomy.

Neurocombinatorial models more generally assume that creativity ultimately rests on a neurophysiological basis. Thagard and Stewart, for example, argue that the mental representations that are combined in the creative act can be reduced to "patterns of neural activity" (2011: 2). They do not explicitly mention "rules", but, if one wanted to turn their theory into a rule-based model, the rules one would need to invoke would be linked to neurophysiological processes and mechanisms. A note on Boden: even though she ties rules to mental rather than neurophysiological processes, that does not mean that she rejects the idea

that mental representations can be explained in terms of neurophysiological processes. Boden (1970) differentiates between two different ways of describing psychological subjects: on the one side, there are "intentional descriptions" which conceive of mental life in terms of mental representations; on the other side, there are "physical descriptions" which make use of physiological and causal terms. In contrast to Koestler and the proponents of neurocombinatorial accounts, Boden's theory is based on an intentional description. Therefore, mental representations form the foundation of her computational model. Even if she may not reject approaches that rely on physical descriptions *per se*, Boden contends that the "intentional description and the physical description are logically distinct, and are not intertranslatable" (1970: 200).

Csíkszentmihályi's systems model opens up yet another perspective: for him, "[c]reativity […] is a process by which a symbolic domain in the culture is changed" (2013: 8). As Csíkszentmihályi holds that what he calls "domain" is organised by rules, one can also put it like this: creativity changes the rules of a domain in the culture. Alas, Csíkszentmihályi does not say much about the (metaphysical) features of domains. However, there is reason to assume that he does not conceive of "domains" in mental or neurophysiological terms. Rather, – in line with his dictum that "creativity does not happen inside people's heads" (Csíkszentmihályi 2013: 23), – domains (and, correspondingly, their rules) are presented as something that is not internal, but external, to the creative individual. That is why, "[i]n order to function well within the creative system, one must *internalize* the rules of the domain" (1999: 332; emphasis added). Since Csíkszentmihályi's "domains" are neither mental, nor neurophysiological structures, it seems most appropriate to characterise them and their rules in accordance with the sociocultural thrust of his theory as sociocultural constructs.

In line with this socio-constructivist reading, Csíkszentmihályi, when he talks about domains and their rules, sometimes uses terms like "consensus" (2013: 28, 37) and "convention" (1999: 332). More frequently, he links the rules of domains to "acceptance"

(Csíkszentmihályi 1999: 318). Accordingly, to change the domain (and its rules), a novelty needs to be "accepted into the domain" (Csíkszentmihályi 2013: 324). In this respect, the gatekeeping instance of the "field" is critical as it decides about which novelties get incorporated into the domain. The field decides about what is acceptable and shapes the consensus that prevails in a domain: "a domain cannot be changed without the explicit or implicit consent of a field responsible for it" (Csíkszentmihályi 2013: 28). So, the field plays a key role in the social construction of domains and their rules.

Even though the rules of a domain are, to a large extent, the result of the judgements of the field, Csíkszentmihályi emphasises that its rules need to be adopted and accepted by everyone who wants to participate in or contribute to the relevant domain: "A person who wants to make a creative contribution […] must learn the rules […] of the domain" and "believe in the importance of such a domain in order to learn its rules" (2013: 47, 71). Thus, when viewed through the lens of the systems model, Picasso's artworks changed the rules of Western art by being judged favourably by galleries, auction houses, museums, and art collectors (that is, the field of the artworld); this renders Picasso's artworks acceptable within the domain of art.[41]

It has been shown that there are three main types of foundations when it comes to theorising about creativity in terms of rules, namely mental (Boden), neurophysiological (Koestler), and socially constructed (Csíkszentmihályi) frames of reference. This raises the question of how *we* should conceive of the "place" within which creativity happens? What is the most adequate foundation or frame of reference when thinking about creativity in terms of rules?

Creativity is not the only phenomenon that is repeatedly theorised about with explicit or implicit reference to rules. In modern philosophy, there is also a strong tendency to think about

[41] Similar to Csíkszentmihályi, Amabile suggests a "consensual definition of creativity" according to which a "product or response is creative to the extent that appropriate observers independently agree it is creative" (1983: 359; also see 1996: 33).

language in terms of rules. Kant and Frege can be regarded as prominent precursors of that vision (Brandom 1994: 7–15). However, arguably the most influential reflection on (linguistic) rules in modern philosophy can be found in Ludwig Wittgenstein's *Philosophical Investigations*. Here, Wittgenstein wrestles with the question of how rules guide language and other human practices such as mathematics. We use language in a way that goes beyond the finite number of linguistic stimuli received as a child. Often, we can understand linguistic utterances we have never heard before. For Wittgenstein, we are capable of doing these things because we are able to (correctly) apply the (relevant) rules of language. Analogously, we are capable of adding integers we have never added before. Probably, you have never in your life been confronted with the task of adding 14,785,314,757 and 2. Nevertheless, you can easily calculate the correct answer.

For Wittgenstein, this raises the question of what it is that determines that the rules of arithmetic are correctly applied when one states that "17,785,314,759", rather than, say, "17,785,314,758" or "42", is the right answer when confronted with "14,785,314,757 + 2". What determines, to use Wittgenstein's example, that "1000, 1002, 1004, 1006", rather than "1000, 1004, 1008, 1012", is the correct continuation of the series "+2" (PI: §185)? What determines that, when following the semantic rules of the English language, the term "orange" is correctly applied to the colour of a tangerine or orange, but incorrectly applied to the colour of a blueberry or a cherry? Wittgenstein scrutinizes various approaches to answering the underlying question of what it is that constitutes correctness when it comes to applying a rule. These approaches can be understood as "the first steps of a search for an account of meaning" (Goldfarb 1985: 488). So, whatever it is that fixes the correct application of such rules promises to be the most adequate foundation (or frame of reference) on which to build a theory of linguistic meaning (and similar rule-governed or rule-related phenomena).

As noted, we can often tell if a rule is being followed or not, even in cases that go beyond what we have experienced in the past. The same often holds with respect to painting and other

arts; towards the end of *Philosophical Investigations*, Wittgenstein even encourages us to "[c]ompare a concept with a style of painting" (PI: 230). Many people can, for example, recognise that a painting was painted by Van Gogh even if they have not seen this particular painting before, but only other paintings by the artist. In other words: a Van Gogh painting follows certain rules that allow us to identify it as a "Van Gogh" (unless it is a copy or forgery) – even if we have not been exposed to this specific Van Gogh painting before. Similar things can be said about new scientific texts, theories, or propositions: usually, we only accept them as "scientific" if they follow certain rules, e.g., if they manifest certain methodological standards and principles and are in line with common knowledge of the relevant scientific domain.

On this basis, even though Wittgenstein himself does not explicitly raise these questions, one can formulate the problem we focus on also in terms of questions that are more relevant to our investigation: what determines, for instance, that the rules of the Bauhaus style are correctly applied when a piece of design or architecture features clean-lined shapes, but incorrectly applied when it features an abundance of ornamentation? What determines that, in order to correctly follow or apply the rules of Cubism, a painting needs to conflate a multitude of perspectives? What is it that determines that, in order to follow the rules of modern physics, one must treat the speed of light as a constant? Whatever it is that settles the matter of correctness with regards to such rules can then serve as a firm foundation when discussing creativity in terms of rules.

With respect to the rule about the constancy of light speed, one might be tempted to respond that physicists today treat the speed of light as a constant because this follows from several experiments and measurements that led to the same results – most prominently, those conducted by Fizeau (1849) and Foucault (1862) – and is crucial to the theory of special relativity. Even though this response might be relevant to explaining why this rule is in place in modern physics, it does not answer the question we are interested in. For we are dealing with

the task of identifying the foundational structure or frame of reference that determines what a rule demands in order to be applied correctly – irrespective of why this specific rule exists within the relevant frame of reference. We do not try to find out what justifies the existence of certain rules (which is specific to the particular rule in question and cannot be answered in a general manner), but what kind of structure it is that makes the existence of rules possible in the first place.

We mentioned several examples of "rules", without spelling out what we mean by that term. In fact, Wittgenstein himself does not define or specify in abstract terms what he means when he talks about "rules". However, I think he uses the term "rule" roughly along the following lines: rules are, positively conceived, recurring patterns (that recur, for example, in the context of certain domains of human life such as art or science) that somehow guide human thought, action, behaviour, and understanding (in these domains) and render things meaningful (in the broadest possible sense). This characterisation of rules is in line with how the term is used in the context of the rule-based theories of creativity under scrutiny. Rules are "recurring" – that means that they require repetition and regularity. Or as Wittgenstein puts it: "The use of the word 'rule' and the use of the word 'same' are interwoven" (PI: §225). Yet, repetition and regularity are not sufficient for something to be a rule. Brandom calls the view according to which *every* regularity constitutes a rule "Regularism" (1994: 26). But this view is problematic as it fails to distinguish between what is done and what ought to be done.

Moreover, for Wittgenstein, rules are connected to meaning and intelligibility (broadly conceived). This is obvious as far as the rules of language are concerned. If someone does not follow the rules of language, it is likely that we will not understand that person. But even beyond that, rules render things meaningful (at least to those who are familiar with them). For example: to those who are familiar with the rules of twelve-tone composition, the musical works of Arnold Schönberg, Alban Berg, and Anton Webern often appear meaningful and intelligible, while, for the uninitiated, they tend to be nothing more than meaningless noise. Likewise,

without the rules of abstract art, the works of Mark Rothko or Joan Mitchell would lack the meaning many of those familiar with the rules of abstract art "see" in them. Similarly, if a "scientific" theory violates the rules of the relevant discipline (e.g., by flouting basic scientific principles or standard assumptions and common knowledge of the field without elaborating on it), the relevant scientific community will most likely not know what to do with it (apart from rejecting it). Some might object that some literary styles such as Postmodernism or Dada literature play with the limits of meaning and, thus, feature, as one might argue, the rule to transcending intelligibility. Yet, one can respond that it is this very rule that renders their unintelligibility intelligible, that allows us to make sense of their transgressions of sense. So, the cases of Postmodernism or Dada do not undermine, but confirm, that rules tend to render things meaningful.

Related to the matter of meaning and intelligibility, rules are often linked to the ascription of value. That is why a scientific theory that is not in accord with the rules of the relevant discipline is likely be dismissed by the scientific community. Likewise, without the rules of twelve-tone music and abstract art, the works that belong to these artistic styles would not be perceived as valuable. Yet, that does not mean that, as long as something is in line with certain rules, it will necessarily be perceived as valuable. A malevolent insult, even if it accords with the rules of language, will (in most circumstances) not be valued. Moreover, there are people who, even though they are familiar with the rules of twelve-tone music, still dislike compositions that follow them. Nonetheless, in many cases, the existence of rules provides the condition of possibility (rather than necessity) for something to be perceived as valuable.

We can now return to the question of what it is that constitutes the correct application of a rule. As we have seen, regularity alone does not do the job. In *Philosophical Investigations*, Wittgenstein attends to various promising explanations of rule-following that propose different foundational entities that are supposed to determine how a rule is applied correctly. These explanatory attempts are "reductionistic" as they try to explain what counts as a correct

application of a rule in terms of more basic entities or mechanisms. More technically speaking, they attempt to reduce facts about the correct application of rules to facts about the foundational entities they invoke. However, Wittgenstein demonstrates that the reductionist explanations he touches on in his remarks on rule-following exhibit serious shortcomings. By drawing on Wittgenstein's criticisms and the philosophical literature they have inspired, I will show that the accounts of rule-following implicit in the theories developed by Boden, Koestler, and Csíkszentmihályi run into similar difficulties.

Wittgenstein's remarks on rule-following, which extend over more than a hundred aphorisms in *Philosophical Investigations*, do not follow a straightforward and concise argumentative structure; often they are sketchy and vague, sometimes even downright obscure (which also seems to be due to the complicated editorial history of the text). For that reason, they have been interpreted in various and, in part, conflicting ways. However, we do not claim that the following reflections represent an accurate depiction of what Wittgenstein himself thought about rule-following. As highlighted in the Introduction, our main concern is not an interpretative, but an argumentative or systematic one.

(a) Against Mentalistic Explanations of Rule-Following

In his discussion of rules, Wittgenstein pays particular attention to the idea that the correct application of a (linguistic or mathematical) rule is fixed by what he calls "interpretations". If this idea is true, facts about the correct application of a rule can be explained in terms of more basic facts about "interpretations". One of the theories Wittgenstein seems to address when he discusses interpretation-based explanations of rule-following is the mentalistic account of meaning offered by C. K. Ogden and I. A. Richards in their 1923 book *The Meaning of Meaning* which Wittgenstein called in a letter to Russell "a miserable book" (quoted from Monk 1990: 214). However, Wittgenstein's criticism is not limited to this theory, but makes a much more general anti-mentalistic point. For "[t]alk of the 'interpretation' of a rule is merely a place-

holder for some particular theory", for example, that the correct application of a rule is fixed by "an 'idea' of the rule present to mind" or, as "a modern cognitivist might claim[, …] an internal representation token" (Proudfoot 2004a: 364). Thus, Wittgenstein's objection to interpretation-based accounts of rule-following extends to mentalistic "explanations that centre on internal representations" (Proudfoot 2004a: 365) more generally. In brief, Wittgenstein attacks the idea that the correct application of a rule is determined by any kind of mental representation.

If the mentalistic picture is accurate, the correct application of a rule is determined by a mental representation or "interpretation" that specifies how to correctly apply the rule in question. However, as shown by Saul Kripke (1982) in great detail, the mental representations or "interpretations" that are supposed to determine the correct application of a rule are themselves open to different interpretations.[42] Introducing a further mental representation or "interpretation" to determine how to interpret mental representations or "interpretations" only pushes the issue one stage back: "we give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it" (PI: §201). In other words: the attempt to explain correctness in the context of rule-following by appealing to mental representations leads to an infinite regress (also see Kripke 1982: 17). So, "any interpretation still hangs in the air along with what it interprets [or attempts to explain], and cannot give it any support" (PI: §198). Hence, one can conclude with Wittgenstein that mental representations or "[i]nterpretations by themselves do not determine meaning" (PI: §198) or, more broadly conceived, the correct application of a rule.

Wittgenstein's critique shows that facts about rules cannot be reductively explained in terms of facts about mental representations. The mental representations or mental rules that structure mentalistic frames of reference such as Boden's "conceptual spaces" are also open to

[42] Kripke (1982: 52f.) convincingly shows that "interpretations" or mental representations are not self-interpreting by highlighting that "interpretations" and mental representations are finite entities and, as such, incapable of determining the application of a rule in infinitely many cases. As a self-interpreting entity must be capable of achieving the latter, "interpretations" and mental representations are not self-interpreting.

different interpretations. Thus, the regress objection also applies to the conception of rule-following implicit in Boden's theory as well as other mentalistic models that rely on mental representations. To avoid potential misunderstandings: I do not presuppose or argue in favour of the non-existence of mental representations.[43] I also do not claim that creative people (or people more generally) do not have any mental representations of certain (aspects of) domains. I only say that what counts as the correct application of a rule is not determined by such mental representations. Therefore, we better desist from using mentalistic frames of reference such as "conceptual spaces" or "mental spaces" as the foundational level of investigation when thinking about creativity in terms of rules.

(b) Against Neurodispositionalist Explanations of Rule-Following

Wittgenstein's regress objection is akin to the so-called "homunculus objection" which has been used to attack mentalistic approaches in the philosophy of mind.[44] Applied to the problem at hand, the homunculus objection criticises mentalistic theories for positing a homunculus inside a rule-follower's head that chooses the right mental representations that specify the correct application of this rule. In doing so, the argument goes, such theories kick off an endless cascade of homunculi. However, some theorists have argued that appeals to homunculi or mental representations do not necessarily lead to an *infinite* regress, but – if rightly conceived – only to a *finite* regress. Daniel Dennett, for example, suggests analysing each homunculus "into *smaller* homunculi, but, more important, into *less clever* homunculi" (1978: 80; also see 2014: 91f.). At some point, he argues, the homunculi will be so stupid that they are equivalent to purely physical mechanisms which do not require further "interpretation". On this basic level, as the argument goes, the regress halts (also see Fodor 1968). Since Boden accepts that – under

[43] Ramsey et al. (1990), Ramsey (1997), and Ramsey (2007), for example, argue for the non-existence of mental representations. Similar eliminativist critiques of specific mental states can be found in Rorty (1965), Churchland (1981), or Stich (1983).
[44] See, for example, Attneave (1960) or Dennett (1969).

an alternative, physical description – "the features of intentionality may be explained by a purely causal account, in the sense that they may be shown to be totally dependent upon physical processes" (1970: 200), this approach does not seem incompatible with hers. Dennett himself invokes homunculi not to develop a theory of rule-following, but to explain how the brain works. However, for defenders of Boden, it may seem like a plausible strategy to update her take on rule-following. The same is true of Fauconnier and Turner who are open to the idea that the mental representations their theory rests on have a neurophysiological and, therefore, physical basis. However, by introducing a committee of increasingly simple homunculi to solve the regress problem, these theories abandon mental representations as their foundational unit of inquiry and introduce a new one, namely physical, more specifically, neurophysiological processes. That is the same frame of reference Koestler's theory rests on.

As we have seen, Koestler reduces the code of rules that governs matrices to "structures in the nervous system" (1964: 631). In other words, he reduces facts about rules to facts about neurophysiological structures and processes. Neurophysiological processes are physical processes and, as such, part of the causal structure of the natural world. Therefore, neurophysiological mechanisms causally dispose us to behave in certain ways. In accordance with that, Koestler remarks that "the rules of [a] newly learnt game become soon automatized to such an extent that it becomes increasingly difficult to go against them" (1964: 639). Once they are fully mastered, the "rules of behaviour which govern established routines", that is, matrices, "function automatically" (Koestler 1964: 639). In other words, Koestler reduces facts about rules to facts about neurophysiological dispositions. For him, when asked to continue the series of "+2", neurophysiological structures dispose us to say "1000, 1002, 1004, 1006", rather than "1000, 1004, 1008, 1012". Likewise, neurophysiological dispositions that constitute the rules of the matrix of the *Bildungsroman* make me correctly apply the rules of this literary genre. Due to their neurophysiological basis, the rules of matrices are "fixed stable rules which, once switched on, automatically govern the thinking routine" (1964: 639).

However, this neurophysiological – or, if you will, "neurodispositionalist" – take on rule-following is also highly problematic. In his seminal work on rule-following, Kripke also formulated influential objections to dispositionalist accounts. The following points of criticism are variations of two of the problems raised by Kripke (1982: 23-37).[45] First, the neurodispositionalist approach fails to account for the normativity of rules: "Rules are things that tell us what we ought to do; dispositions are simply facts about what we would do. To think that facts about the rules we follow can be derived from facts about our dispositions is to try, illegitimately, to derive an 'ought' from an 'is'" (Holton 2001: 9432).

Secondly, imagine that a neurophysiological disposition in my brain causes me to respond "1000, 1004, 1008, 1012" when queried to continue the series "+2". According to the neurodispositionalist account, this would be the correct response as it is in line with a neurophysiological disposition I have. This conclusion is, of course, problematic. Since the neurodispositionalist account identifies the correct application of a rule with how one's neurophysiological structures cause or dispose one to act, it fails to account for the possibility that someone applies a rule in an incorrect manner. In other words, it undermines the possibility of someone making an error. In doing so, it rules out the possibility of breaking a rule. What is more, it even undermines the very idea of a rule – for it is part of the idea of a rule that is possible to break it. Besides that, as outlined in section 3.2, a satisfactory take on creativity should account for the circumstance that some instances of creativity change existing frames of reference. In other words, creative achievements of this kind change the rules of a given frame of reference. Changing the rules of a frame of reference often requires breaking its rules. Consequently, an adequate account of creativity should accommodate the possibility of doing so.

[45] Wittgenstein's own remarks on causal, disposition-based takes on rule-following are quite vague and fragmented, but it becomes clear that he is highly critical of such approaches (see, e.g., PI: §§149, 220).

Many attempts have been made to refute Kripke's anti-dispositionalist criticisms. However, even philosophers with dispositionalist tendencies admit that "none fully succeeded" (Warren 2020: 257). The latest outline of a rectified dispositionalist account that aims to be immune to the points of criticism presented above was developed by Warren (2020). Let us focus on his response to the problem of the impossibility of error. To overcome this problem, he proposes that we should appeal to dispositions in *normal situations*. In order to avoid circular specifications about which situations qualify as "normal" (as criticised by Kripke 1982: 30f.), he defines "normal situations" as "those in which neither external nor internal factors are interfering with [an individual's] general cognitive functioning" (Warren 2020: 270). When the rule-follower is, for example, drunk or extremely tired, the situation is anormal due to internal factors. A situation "with mind-altering chemicals permeating the air isn't normal" (Warren 2020: 270) because of external factors. In other words, only in cases of "clearheadedness" (Warren 2020: 274), the dispositions of an individual are decisive. At the same time, Warren accepts that, even in normal situations (that is, when the rule-follower is clearheaded), it is not impossible to make mistakes. So, he assumes that it is suffices if one is, at least, "in the *vast majority* of normal situations" (Warren 2020: 271) disposed to apply a rule correctly. Warren adds that a person's disposition, in order to determine correctness, must also be "*stable*" which means that a "sequence of checks and rechecks tends to produce the same answer [or application of a rule] over and over again" (2020: 271). Warren does not further elaborate on how such checks and rechecks to "screen out errors" (2020: 272) are being executed *in concreto*. Obviously, one cannot appeal to one's own performance of applying a rule to check the correctness of one's disposition as this would lead to a vicious circle. Therefore, it seems plausible to assume that, at least in many cases, one's social environment serves as a regulative instance: erroneous dispositions are, for example, rectified if a person is made aware of the fact that she applies a rule incorrectly (since her way of applying is not in line with the dispositions of the individuals that notice her mistake). Or the person notices that all other people she

encounters apply a rule in a way that varies from the way she applies it so that she "corrects" the relevant disposition. To sum up: according to Warren, facts about what counts as the correct application of a rule can be reduced to facts about how an individual is stably disposed to apply a rule in the vast majority of normal situations (in which the individual is clearheaded).

However, Warren's updated version of dispositionalism is also not unproblematic. Let us assume that a person is, in in the overwhelming majority of normal situations, stably disposed to misuse a certain rule. Imagine, for example, a person that is, in practically all normal situations of clearheadedness, stably disposed to refer to Impressionistic paintings, artists, and techniques as "Cubist". This is not due to drunkenness or mind-altering chemicals, but simply because the person has internalised an incorrect way of using the terms "Cubist", "Cubism", etc. One might be tempted to respond that the person cannot be *stably* disposed to make this mistake as other people will correct the mistake so that the disposition of the person will be rectified. However, it is possible to imagine that all people in the person's community make the same mistake.

Imagine, for example, that there is a copy of Renoir's *Luncheon of the Boating Party* in the townhall of the relevant community. Let us assume that the community has no access to art-historical books, but only an album that features famous artworks from different periods without accompanying commentary. Imagine that a foreigner told the members of the community (which is very isolated from other communities), for whatever reason, that the painting in the townhall is a "Cubist" artwork. Therefore, all members of the community refer to the picture in the townhall and all the pictures in the uncommented album that resemble it stylistically as "Cubist". Accordingly, a person that is part of this community is – in the vast majority of normal, clearheaded situations – disposed to call Impressionist paintings "Cubist". And the person is *stably* disposed to do so for checks and rechecks will not change this disposition as all the other members of the community use the term "Cubist" in the same way. So, since both the conditions of clearheadedness and stability are satisfied in this scenario, Warren's

dispositionalist account implies that the member of this community applies the terms "Cubist" correctly. However, we would still want to say that the person makes a mistake.[46]

But back to Koestler: A defender of his account might highlight that he does not only speak of fixed "rules" that are reducible to patterns of the central nervous system and "govern established routines and function automatically" (1964: 638), but also of flexible "strategies" which are "guided by environmental pointers" (1964: 38). However, the flexibility of such strategies is limited by the rules of the matrix. In Koestler's own words: "The game of chess allows you a vast number of strategic choices among the moves permitted by the rules, but there is a limit to them" (2009: 253). So, there cannot be a strategy to break the rules. Others might point to the circumstance that Koestler holds that some rule must be learnt (for example, the rules of chess) and take this as evidence that his theory implies that there are more basic rules which are independent of neurophysiological dispositions. They might continue to argue that these more basic, non-dispositional rules (whatever their positive metaphysical features may be) make it possible to accommodate the possibility of applying a rule incorrectly. However, this response only confirms that Koestler's matrices represent an inadequate foundation for a rule-based account of creativity.

The criticism levelled against Koestler's take on rule-following can also be directed against other models of creativity that rest on a neurophysiological foundation. This includes a "naturalised" version of Boden's account that introduces a regress-stopping army of increasingly stupid homunculi which are, at their most basic level, nothing but physical, that is, causal mechanisms. We do not make any claim here about whether mental representations can be analysed in terms of a cascade of increasingly stupid homunculi and, thus, be explained in terms of neurophysiological processes. According to Dennett, "bottom-up, neuroscientifically inspired, homuncular functionalism" is a "model of how the brain works" (2014: 95). And

[46] For another critique of Warren's approach, see Miller & Sultanescu (2022).

maybe some form of it is an accurate model of how the brain works. Yet, that does not mean that it is also an accurate model of rule-following. For, when applied to the issue of rule-following, the most basic homunculi are still neurophysiological processes. And it remains unclear how these can account for the normativity of rules.

(c) Against Socio-Constructivist Explanations of Rule-Following

It has been shown that facts about rules cannot be reduced to facts about mental representations or facts about neurophysiological mechanisms and dispositions. Yet, there is still Csíkszentmihályi's theory which introduces "domains" as its foundational unit of inquiry. Even though he states that a "person who wants to make a creative contribution […] must also reproduce [the relevant] system within his or her mind" (2013: 37), the mental replication (or mental representation) of the system or domain does not settle what qualifies as the correct application of a rule. Rather, whatever it is that is internalised or replicated determines how a rule is applied correctly. As outlined above, Csíkszentmihályi seems to paint a socio-constructivist picture of rule-following according to which the rules that govern a certain domain are a matter of consensus, convention, and acceptance within the community of the respective domain. Consequently, it seems plausible to assume that, according to his model, facts about rules are determined by facts about what is agreed upon within the relevant community.

Some philosophers proposed a similar approach in the context of the debate about linguistic and mathematical rule-following.[47] According to this view, the correct application of a (linguistic or mathematical) rule is determined by what a community has agreed to be the correct application; facts about rules can be reduced to facts that are constructed through (implicit or explicit) social consensus. Unlike the dispositionalist account of rule-following,

[47] See, for example, Dummett (1978), Wright (1980), and Malcolm (1986). Moreover, Kripke's (1982) take on rule-following is often interpreted along similar constructivist lines.

this socio-constructivist perspective manages to account for the possibility of an individual applying a rule incorrectly: individuals flout or break rules if they do not comply with what the community has agreed on or accepts as correct.

However, the socio-constructivist take on rule-following also exhibits some shortcomings. To expose these, let us go back to the thought experiment that was introduced to criticise Warren's dispositionalist account. Let us assume that the foreigner who is a French speaker tells another lie to the members of the isolated community that do not speak French: he tells them that, in his language, "maison" means "forest". Let us now imagine that, due to an apocalypse-like event, all humans go extinct, except for the members of the isolated community and the foreigner and his family. As a result, the members of the isolated community greatly outnumber the French-speakers on earth. This means that the vast majority of people on the planet believe and agree that "maison" means "forest". Proponents of the socio-constructivist account have to conclude that, in this case, "maison" actually means "forest". Accordingly, the socio-constructivist picture implies that the foreigner and his family – that is, actual French speakers –use the term "maison" incorrectly. Analogously, according to the socio-constructivist view, the members of the isolated community are correct in saying that "maison" means "forest". However, we would say that the community is wrong in saying that "maison" means "forest", as it has been deceived. Likewise, we would say that the foreigner and his family are, as French speakers, correct in how they use the word "maison". Yet, the socio-constructivist account does not allow us to acknowledge this.[48] Hence, the attempt to explain the correct application of rules in terms of communal agreement and acceptance is not successful. We can agree with Wittgenstein that "what is [correct] and what is false […] is not agreement in [shared] opinions [or beliefs]" (PI: §241).[49]

[48] For another critique of the socio-constructivist approach, see Boghossian (1990: 535f.).

[49] Anscombe's actual translation speaks of "what is true and what is false". However, in the German original, Wittgenstein writes: "Richtig und falsch [...] ist keine Übereinstimmung der Meinungen" (Wittgenstein 1975: §241). The German adjective "richtig" bears normative connotations so that "correct" represents a more adequate translation than "true".

However, defenders of the systems theory might point to the fact that, according to Csíkszentmihályi, the field determines what qualifies as acceptable. Its judgements form the consensus in the domain. So, they might continue to argue that the account of rule-following underlying the systems theory should be characterised differently, along the following lines: the correct application of a rule is determined by what the field accepts as correct; facts about rules derive from facts about the judgements of the field. However, in the above-mentioned thought experiment, the people who could possibly represent the field within the isolated community (e.g., the community members who were told by the foreigner that "maison" means "forest" or those who are considered to have the most extensive knowledge of language or forests) are under the same deception as all other members of the community.

One might respond that the "field" in this example does not qualify as a field as it is deluded; therefore, we should demand that only non-deluded individuals represent the field. Let us assume that the family members of the foreigner also pass away. In this case, in our thought experiment, the field consists solely of the foreigner. However, imagine that the foreigner returns to the community many decades after his first visit and tells its members that "maison" actually does not mean "forest". It is easily imaginable that the members of the community will not believe him – maybe because they forgot who he is or, if they recognise him, because they mistrust someone who admits that he once lied to them. As a result, the community consensus about what the word "maison" means does not change. In that case, the foreigner does not represent the field as conceived by systems theory. For, according to Csíkszentmihályi, the field shapes the consensus among the community. Since the community does not trust the foreigner, he does not have an influence on the community consensus and, thus, does not qualify as the field. As a result, appealing to the field does not solve the problem.

(d) Against Platonist Explanations of Rule-Following

Lastly, there is another venerable approach to rule-following: Platonism. Platonic metaphysics assumes that there are perfect, universal entities (Platonic Ideas or Forms) that are located in a realm that is separate from the natural world and the world of the mind. So, according to the Platonist account of rule-following, the correct application of a rule is determined by these objective, supernatural, and mind-independent metaphysical entities. Such a Platonist take on rule-following was, for example, proposed by Frege (1956: 302).

In today's creativity studies, Platonic metaphysics does not serve as a major point of reference. In large part, this seems to be due to the fact that, in line with the Paradigm 2 understanding of creativity, current creativity research aims to explain *human* creativity. Platonic metaphysics, however, as a main representative of Paradigm 1, undermines the idea of genuine human creativity. However, even though Platonic approaches to creativity are, for that reason, not very popular today, that does not mean that they do not exist.

Recently, Taliaferro and Varie proposed a Platonic approach to creativity by explaining creativity in terms of metaphysical "states of affairs" which are equivalent to Plato's "Forms" or "Ideas" and pre-exist all creative achievements: "all […] creative works existed as abstracta prior to their obtaining. That is, before Picasso painted *Guernica* or Mozart composed *Don Giovanni*, both existed as states of affairs that could be obtained but were not. The role of the artist, then, was to make the states of affairs obtain" (2018: 145f.). These metaphysical states of affairs serve as the foundational unit of this Platonic theory of creativity. The Platonic theory of creativity can, among other things, be criticised for de-historicising and de-contextualising creative achievements, and for its distinctive lack of parsimony, as it not only assumes that "all past, present, and future creative objects" and "each phase of a given creative object", but also "all possible created objects that will never obtain […] exist as states of affairs" (Taliaferro & Varie 2018: 146).

Furthermore, the Platonic take on rule-following according to which eternal Platonic objects (such as Taliaferro's and Varie's "states of affairs") determine the correct application of a rule is problematic. For it is unclear how such abstract, supernatural entities can be accessed by human rule-followers. As Platonic objects (states of affairs) exist outside time and space, they exist outside the cause-and-effect relations of the physical universe. Therefore, positing a causal relation between Platonic objects and the behaviour of rule-followers seems futile: "whatever such objects might be exactly, […] as abstract objects they cannot causally interact with individuals in space and time and thus cannot explain the acts of individuals that are in space and time" (Haase 2012: 240). Platonists might accept that the abstract objects of the Platonic heaven (states of affairs) do not causally interact with rule-followers but contend that the human mind has access to – can represent – these supernatural objects. Now, however, there is a gap between the Platonic object (state of affairs) that determines the correct application of the rule and the mental representation of it. We are confronted with the issue of bridging this gap, that is, the question of what determines how to correctly interpret or mentally represent the relevant Platonic objects (states of affairs). Invoking further internal representations or "interpretations" would not solve this issue as it would unleash an infinite regress.

Platonists might respond by claiming that Platonic objects (states of affairs) are self-interpreting rules. This, of course, raises the question of whether such things as self-interpreting rules really exist or should be dismissed as a "platonist mythology" (McGinn 2013: 92). Yet, even if one accepts, first, that there are such things as self-interpreting rules and, secondly, that Platonic objects represent such rules, it seems vain to eliminate any kind of involvement of the human mind. Or as Kripke puts it: "Platonic objects may be self-interpreting, or rather, they may need no interpretation; but ultimately there must be some mental entity involved" (1982: 54) that reintroduces the infinite regress problem. For Kripke, that is because human minds (and their internal representations) are finite, whereas rules, if understood as Platonic entities, are supposed to apply to infinitely many cases. (Wittgenstein himself compares Platonic objects

with “rails invisibly laid to infinity” which “correspond to the unlimited application of a rule” (PI: §218).) Independent from Kripke’s take on the matter, there is another reason to assume that, even if Platonic objects are self-interpreting rules, there needs to be some kind of mental mediation. For otherwise it seems, again, impossible to explain and accommodate the possibility of making a mistake, that is, of applying a rule incorrectly.

The Platonists are, therefore, confronted with a dilemma: they either need to abandon the possibility of making a mistake in which case they face the same issue as the proponents of the dispositionalist account; or they need to invoke further mental representations or meta-rules to bridge the gap between the mind of the rule-follower and the relevant Platonic objects (states of affairs) which leads to an infinite regress. Consequently, as it remains unclear how human rule-followers can access Platonic objects (states of affairs), the Platonist picture of rule-following is, in Wittgenstein’s words, not more than “a mythological description of the use of a rule” (PI: §221). Hence, Platonic objects like Taliaferro’s and Varie’s “states of affairs” cannot serve as the foundation of a rule-based theory of creativity.

3.4 Towards a Practice-Based Account of Creativity

As we have seen, rule-following cannot be reductively explained in terms of mental representations, neurophysiological processes, communal agreement, or universal Platonic objects. Facts about the rules that govern our practices cannot be reduced to more basic facts about these kinds of entities. In the face of the problems reductionist explanations of rule-following generate, philosophers like McDowell (1984; 1992; 2009), Stroud (2012; 2000), and, arguably, Diamond (1989) claim that Wittgenstein does not try to explain rule-following in terms of anything else. According to this reading, the aim of his rule-following discussion is to therapeutically free us from the misguided urge to reduce facts about rules to more basic facts. Instead, Wittgenstein provides an anti-reductionist or “deflationary” take on rule-following: “Wittgenstein takes facts about rules and standards of correctness as basic and irreducible; he

does not attempt to construct those facts from more basic non-normative facts" (Child 2011: 130). So, instead of trying to explain rule-following in a reductionistic manner, we should view rules as embedded in practices: "Wittgenstein's point is that we have to situate our conception of meaning and understanding [or, technically speaking, of rule-following] within a framework of communal practices" (McDowell 1984: 342). In Wittgenstein's own words: "'obeying a rule' is a practice" (PI: §202). Or specifically with regards to linguistic or, more precisely, semantic rule-following: "*Practice* gives the words their sense" (Wittgenstein 1980a: 85e). According to the anti-reductionist, deflationary take on rule-following, rules (that is, recurring patterns that are meaningful) and practices are mutually dependent: there cannot be a rule without a practice, nor a practice without rules.

Due to the shortcomings of the reductionist attempts to explain rule-following that have been discussed above, I also embrace an anti-reductionist or deflationary approach to rule-following, but one with significant qualifications. Deflationary interpretations of Wittgenstein's remarks on rules sometimes read as if the later Wittgenstein has (or claims to have) proven, once and for all, that facts about rules cannot be reduced to more basic facts and are, therefore, *sui generis*. However, I think that such a view is misguided. First, such a view is incompatible with Wittgenstein's metaphilosophy. The claim that facts about rules are *sui generis* is a metaphysical thesis. So, by holding such a view one falls back on the metaphysical mode of doing philosophy the later Wittgenstein aims to overcome.

Secondly, Wittgenstein's reflections on rule-following show that the most promising attempts to reduce facts about rules to facts "outside" of the practices they are embedded in gets us entangled in infinite regresses, paradoxes, and other problems. Therefore, one had better refrain from such reductionist endeavours. But this does not *absolutely* rule out the possibility of a reductive explanation of practices and, thus, the rules that are embedded in them. Maybe we have just overlooked the right foundational entity so far. Or maybe it is beyond our cognitive powers to identify it. Therefore, I am agnostic as to whether it is possible to actually prove that

a reductionist explanation of rule-following is impossible and that facts about rules are *sui generis*. Given this agnosticism about the possibility of reductive explanations of rules, we adopt an anti-reductionist perspective on rule-following for pragmatic rather than substantive ontological reasons. That is, because we do not know what the explanans could be and if there is one in the first place, it would be better to desist from theorising about rule-following in reductionist terms and, rather, regard rules as embedded in practice.

Practices are not metaphysical entities, hidden in a transcendent realm. Practices are part of our lives. Therefore, those who are familiar with a certain practice normally have no problem identifying correct and incorrect applications of the rules that are part of this practice – even if they cannot explain why a certain rule has or has not been applied correctly in a specific situation. Likewise, we normally have no problem following the rules of the practices we are familiar with. Even if there is a complicated, for humans probably unfathomable, reductionist explanation of rule-following, this would not change anything. For we do not need such a reductionist explanation in order to be able to follow rules and identify misuses of rules. We "obey the rule[s] *blindly*" (PI: §219). So, Wittgenstein's treatment of rule-following, as we interpret it, is akin to his take on whether every concept needs to be definable in terms of necessary and jointly sufficient conditions. Just as the late Wittgenstein abstains from making any metaphysical claims about this subject matter, he does not make any metaphysical claims about whether there is or is not a reductive explanation of rule-following. Instead, he merely reminds us of the irrelevance of these things when it comes to using language and following rules: such definitional questions "never troubled you before when you used the word 'game'" (PI: §68). The same holds true for questions about rule-following: rules become only problematic if we detach them from the practices they are embedded in.

In line with that, practice does not represent a metaphysical foundation or ground. Instead, practice creates, as one might frame it, a "semblance of ground" (Moi 2017: 29), a "thin net over the abyss" (Cavell 1979: 178). Even if practice is a ground without ground –

"groundless ground" (Braver 2012) – it is still a ground. It is the "firmest" ground available to us. For, at some point, "[g]iving grounds[…] comes to an end" (Wittgenstein 2016b: §204) – and all that is left to say is: "This is simply what I do" (PI: §217). Therefore, practice is the most suitable point of reference when thinking about rule-following. Accordingly, in the following analysis of "creativity", normative practices will serve as the basic unit of inquiry.

However, how to conceive of rules as embedded in practice? According to Peter Hacker's influential interpretation, rules *govern* our practices, particularly our linguistic practice. Yet, Hacker emphasises that the rules that govern our practices are not hidden metaphysical entities; rather, they are based on how we ordinarily use words (or, more generally, how we ordinarily do things). Rules capture "how the relevant words are generally used" (Hacker 2007: 100). According to Hacker and like-minded exegetes, these "familiar, accepted rules" (2013: 156) govern practice in such a way that nonsense arises whenever one transgresses these rules. Accordingly, for Hacker, "[t]he aim of philosophy is to stop people from transgressing the bounds of sense" (Hacker 2007: 100), i.e., from breaking the rules. Quite fittingly, Hacker's Wittgenstein has been described as a "nonsense policeman" (Macarthur 2014: 6). However, Hacker's take on rules is too restrictive and rigid. While Wittgenstein himself considers languages and practices "dynamic phenomena" (Eriksen 2020: 81), Hacker's reading neglects the flexibility of language. An adequate account of rule-following should resist the temptation "to imagine rules as containing their own applications in a way that forecloses the future [… and to] think of rules as anticipating reality" as if "the pattern of their future uses is predetermined in some peculiar way" (Mulhall 2001: 118). Yet, Hacker's reading succumbs to this temptation. For that reason, his interpretation fails to give a satisfactory account for the fact that our linguistic practice can change – or, more generally, that the mutation of a practice is no anomaly.[50]

[50] Crane (2015: 255) offers a similar critique. It might be replied that, whenever a word is used in a new way, the rules that govern its use are changed to settle a new use for it. Yet, the very *transgressions* of given rules that are

This raises the question how we should conceive of rules as embedded in practice? In contrast to Hacker's account which implies that rules are more fundamental than their practices, a truly anti-reductionist and deflationary approach should assume a "reciprocity" (Taylor 1995: 178) between rules and their practices. When a rule changes, the practice it is part of changes; when a practice changes, its rules – that is, its recurring patterns of meaning – change. When a new pattern comes up and is repeated within a practice in a meaningful way, a new rule emerges. That happened, for instance, when people started to say "you" in contexts where they would have previously said "thou" or when artists started to merge various perspectives in a single picture with the inception of Cubism. This does not mean that we reduce facts about rules to non-normative facts such as facts about use or facts about something being in circulation. For use and being in circulation alone do not establish a rule. Rather, something must *meaningfully* be used or circulated in order to represent a rule. So, in line with the idea of a reciprocity between practice and rules, practices are imbued with meaning (Stroud 2012: 27). (Yet, the fact that a rule is less and less used and, therefore, less and less in circulation, might have the effect that a rule gradually loses its normative power until it dissolves.)

Wittgenstein states that a game is "not everywhere circumscribed by rules" (PI: §68). There are, for example, no rules for "how high one throws the ball in tennis, or how hard; yet tennis is a game for all that and has rules too" (PI: §68). Wittgenstein frequently compares our linguistic practices with games – think of his "language-game" (PI: §7) metaphor. Consequently, he holds that not every aspect of our practices is covered by rules. (Accordingly, Wittgenstein rejects the idea that whenever we speak – or, more broadly, whenever we follow rules – we are "thereby operating a calculus according to definite rules" (PI: §81).) Since not every aspect of a practice is determined or covered by rules, new rules can emerge within

necessary to establish a new use (and, thus, a new rule) are, in Hacker's view, illegitimate and objectionable. Yet, language change is no anomaly, but a common phenomenon. So, it is strange to assume that it is made possible by a nonsense-producing philosophical crime. One might respond that Hacker seems to react to this issue by arguing that conceptual change is made possible through analogy (Baker and Hacker 2005: 177). This, however, leads to another problem: when is a transgression of a rule justified by analogy – and when is it just nonsense?

established practices. The first time somebody said "you" instead of "thou", there was not yet a rule for using the word "you" as the second-person singular. However, when more and more people started using and understanding the term "you" in such a way, such a rule emerged.[51]

Moreover, some rules are more crucial for a practice than others. Rules that play a crucial role in a practice are often rules which are repeated very often within that practice. Such rules are often harder to change, eradicate, or ignore. It is, for example, much harder to change the subject–verb–object structure of English syntax than, say, to change the rules for how to use the term "highfalutin". Similarly, for the current practice of physics, the rules set by relativity theory play a more crucial role than the rules set by string theory. Consequently, it is more problematic to ignore the former rules than the latter. To characterise our linguistic practice, Wittgenstein invokes the metaphor of the "bank of [a] river" that "consists partly of hard rock, subject to no alteration or only to an imperceptible one, partly of sand, which now in one place now in another gets washed away, or deposited" (Wittgenstein 2016b: § 99). Using this metaphor, one can compare the very crucial rules of a practice to the "hard rock" and its less important rules to the "sand". Rules of the latter kind are usually less often used; some rules might be so uncommon or parochial that one might be corrected for using them. Think, for example, of teenagers talking in some peculiar teenage slang to their grandparents. Alternatively, when confronted with a rule that is barely used, people might have a discussion about whether that rule applies (is actually a rule) or not.

The anti-reductionist view that facts about rules cannot be *reduced* to facts outside of the normative practices they are embedded in does not necessarily imply that there are no other, weaker relations between the facts of the former and the latter kind. Child, for example, recently criticised what he calls "the most flat-footed, uncompromisingly anti-reductionist position"

[51] To avoid potential misunderstandings: I do not advocate the position Brandom calls "Regularism" according to which *every* regularity constitutes a rule. According to my perspective on rules, only somewhat established regularities which are part of a practice and intelligible to others form rules. So, according to my approach, rules involve regularity; but the latter is not sufficient for the former.

(2019: 212) which is sometimes attributed to the late Wittgenstein. According to this extreme form of anti-reductionism, any attempt to relate facts about rules to non-normative facts (outside of normative practices) ought to be dismissed. For Child, instead, "Wittgenstein's position […] combines anti-reductionism with a willingness to go beyond merely pleonastic statements about […] rules and rule-following" (2020: 112). I agree with Child that it is possible – and in line with Wittgenstein's later philosophy – to relate facts about rules to non-normative facts (that is, to facts outside of the normative practices the relevant rules are part of) in an informative, but non-reductionistic way. However, I approach the matter from a different angle.

The anti-reductionist, deflationary perspective we adopt for pragmatic reasons treats rules as basic features of our normative practices. Therefore, to relate facts about rules to non-normative facts, we need to examine how practices relate to facts beyond themselves. When we resist the philosophical drive for abstraction and generalisation and attend to actual human practices, we recognise that practices are part of a wider context. This context gives significance to practices. In line with that, when Wittgenstein describes scenarios of "primitive language[s]" or simple "language-game[s]" (PI: §7) in *Philosophical Investigations* to cast light on our language-use, he provides information about the context within which these rudimentary language-games are played. However, not every context facilitates any practice. The practice of fishing, for example, is a pervasive practice among the Inuit, but not among the Tuareg. Thus, we can say that facts about practices are shaped by non-normative facts that are part of the wider context within which practices are situated. Consequently, facts about rules, as they are embedded in practices, are shaped by the facts that shape the practices they are embedded in.

How should we positively characterise the wider context of facts that shape our practices (and, thus, the rules embedded in them)? Wittgenstein sometimes calls the context within which the "practice of the use of language" (PI: §7) and other rule-following activities are situated "form of life" or, in German, "*Lebensform*". Accordingly, he writes that "to imagine a language

means to imagine a form of life" (PI: §19) and that "the speaking of language is part of […] a form of life" (PI: §23). This intertwinement with forms of life is not only restricted to the practice of language. Practices like praying, cooking, science, or painting are all related to a specific *Lebensform*. Wittgenstein makes clear that we should not try to relate what he calls "form of life" to something more basic; in his own words: "What has to be accepted, the given, is—so one could *say—forms of life*" (PI: 226). That, however, does not mean that facts about our practices are *reducible* to facts about forms of life.[52] Rather, a form of life is the backdrop against which our practices and their rules appear meaningful, significant, and purposeful. (More on the relation between practices, rules, and forms of life in a bit.)

Wittgenstein famously states that, "[i]f a lion could talk, we could not understand him" (PI: 225). However, this does not mean that Wittgenstein's concept of a "form of life" is a biologistic concept that is roughly equivalent to "species". Quite to the contrary, many Wittgenstein scholars associate the term exclusively with culture rather than biology – as synonymous with "culture or social formation" (Glock 1996: 125). However, focusing solely on cultural and social constellations makes, as Stanley Cavell rightly objects, "the teaching of the *Investigations* much too[…] conventionalist" (1988: 255). For Wittgenstein himself writes that correctness is "not agreement in [shared] opinions but in form of life" (PI: §241). So, one should not construe the latter in terms of the former. Moreover, such a conventionalist approach runs the risk of reintroducing the problems the socio-constructivist account of rule-following faces.

To challenge solely cultural interpretations of the concept of "form of life", Cavell himself differentiates between two senses of the Wittgensteinian notion, namely a "biological" (1988: 255; 2013: 41) sense, on the one side, and an "anthropological" (1988: 255) or "ethnological" (2013: 41) sense, on the other. For Cavell, "the romance of the hand and its

[52] Nonetheless, the notion of form of life is sometimes described in reductionistic terms as the "'bedrock' which provides the ultimate basis for meaning, use, rules" (Grayling 2001: 122).

opposable thumb […], and of the upright posture and of the eyes set for heaven" (1988: 255) are biological factors or facts that set the human form of life biologically apart from other forms of life. The anthropological or ethnological sense of form of life, on the other hand, is linked to "a conventionalized, or contractual, sense of agreement" (Cavell 1988: 255). In short: Cavell here links the biological sense to "the natural" (2013: 44) and the anthropological or ethnological one to "the social" (2013: 44). Since Cavell argues that an accurate Wittgensteinian take on language must account for both the biological and the anthropological, he rejects the idea that language is *solely* based on conventions. However, the picture of the notion of "form of life" that is sketched by Cavell's distinction is still incomplete. Consider the circumstance that fishing is a pervasive practice in the form of life of the Inuit, but not of the Tuareg. As the Inuit and the Tuareg share the same form of life biologically understood, this circumstance cannot be traced back to a difference in biological facts. At the same time, it seems misguided to account for this circumstance mainly in terms of convention and agreement. Rather, the fact that fishing is part of the form of life of the Inuit, but not of the Tuareg, is primarily due to facts about the environments inhabited by both groups. Yet, when Cavell elaborates on the two dimensions of forms of life he introduces, there is no mention of the environment.[53]

From our deflationary perspective, the best way to approach the notion of "form of life" is by proceeding from our human practices. From this point of view, to attain a better understanding of the human form of life overall as well as the different forms of life it encompasses, we must ask: what kinds of facts shape or impact our human practices? Several (in part interconnected) facts come to mind:

(a) *Facts about the biology and anatomy of our species*: Facts about our human biology and autonomy such as thumb opposability and upright posture have a significant impact on our

[53] A less conventionalist or contractual understanding of the conventional can, arguably, be found in Cavell (2018: 64–100). Yet, due to space limitations, we cannot discuss here if this less conventionalist conception sufficiently accounts for the role of the environment.

practices. Think, for example, about how the human practices of painting, writing, tennis, or sword fighting, and many others depend on us having opposable thumbs. If humans did not have opposable thumbs, these practices would either not exist or have a vastly different form.[54] Or imagine that humans had no feet, very limited eyesight, but a much better sense of smell. In that case, there would be no practice of football and the practices falling under the visual arts would either not exist at all or be very different from ours; however, there would, most probably, be a much greater variety of highly sophisticated olfactory practices.

In *Philosophical Investigations*, one finds similar thoughts. For Wittgenstein, our practices are interwoven with our "natural history" (PI: §25). This does not mean that he attempts to reductively explain the former in terms of the latter. Nonetheless, natural facts have an impact on our practices and their rules: "if anyone believes that certain concepts [or, more generally speaking, certain rules] are absolutely the correct ones[…]—then let him imagine certain very general facts of nature to be different from what we are used to, and the formation of concepts [or rules] different from the usual ones will become intelligible to him" (PI: 230). Similarly, Wittgenstein remarks that "[w]hat we have to mention in order to explain the significance, I mean the importance, of a concept, are often extremely general facts of nature" (PI: §143). In brief, Wittgenstein acknowledges that facts about rules relate to facts about nature. Facts about the biology and anatomy of the human species represent facts of nature; so does the next kind of facts.

(b) *Facts about the physical environment and their affordances*: Practices are often tied to a certain physical environment. The circumstance that the practice of fishing is a pervasive practice among the Inuit, but not among the Tuareg, illustrates this. The notion of "affordances" is helpful in this context: "The *affordances* of the environment are what it *offers* the animal,

[54] One might object that writing and painting do not necessarily require thumb opposability as we could, in principle, use our feet to write and paint. That is true, but if we all used our feet to write and paint, the practices of writing and painting would most likely look very different. We would have different painting styles and most likely different letters.

what it *provides* or *furnishes*" (Gibson 2015: 119). If the environment inhabited by the Innuit did not contain waters, it would not "afford" the practice of fishing. More broadly speaking: if we lived in a world with vastly different affordances, many of our practices would also be different, or would lose their significance.

Wittgenstein himself makes a similar point when he writes that "[t]he procedure of putting a lump of cheese on a balance and fixing the price by the turn of the scale would lose its point if it frequently happened for such lumps to suddenly grow or shrink for no obvious reason" (PI: §142). The affordances of the fictitious world that Wittgenstein describes here are different from ours. In such a world in which things like food changed their weight in an irregular way, the practice of weighing would, as Wittgenstein remarks, lose its significance. (Moreover, one can add that, in this imaginary world, (the rule of) weighing would lose its significance in the context of the practice of commerce.)

(c) *Facts about conventions and agreements*: Even though conventionalist interpretations of Wittgenstein's notion of "form of life" fall short, that does not mean that (more or less arbitrary) conventions and agreements do not shape many human practices. Think, for example, about the practice of chess: its rules (e.g., the rule the rook can only be moved horizontally or vertically, but not diagonally) are based on conventions. Or think about the impact musical tuning systems (which define the pitches or tones used to tune an instrument) have on our musical practices: they predetermine the tones that can be played. Accordingly, alternative tuning systems open up different musical possibilities. Compare, for instance, the twelve-tone equal temperament that dominates Western music to microtonal tunings that are more common in other cultures – and the different musical patterns they facilitate.

(d) *Facts about our desires and volitions*: Besides the above-mentioned factors, facts about our desires and volitions shape our human practices. The practices of science and philosophy are, for example, at least to some extent, connected to the human desire for knowledge – or, as

Nietzsche would more dramatically put it, the "*will to truth*" (2007: 112). That does not mean that these two practices are exclusively linked to the desire for knowledge; they are also related to other human desires, for instance, the desire for the "mastery of nature" (Horkheimer & Adorno 2002: xviii) or the desire for a just society. Our artistic practices, on the other hand, are – among other desires – linked to the desire for aesthetic experiences. Many practices, including agriculture, medicine, or hunting (at least, in hunter-gatherer societies), are more or less directly connected to the will to survive. In short: if we had different desires and volitions, we would most likely have different practices.

This list is not comprehensive, but it captures important aspects of the human form of life. For some practices, some of the above-mentioned factors are more crucial than others (consider, for example, the role conventions play in the practice of chess compared to the practice of swimming.) To see the interplay of factors (a) to (d), consider the human practice of cooking and its rules which are, among other things, shaped by: (a) facts about our bodily requirements for certain nutrients and facts about our digestion; (b) facts about the nourishments available in the environment; (c) cultural conventions and agreements, e.g., to not eat certain animals even though they are edible and inhabit the same environment; (d) the desire to stay alive (that is why we do not use poisonous plants as ingredients) or the desire for good taste.

From our deflationary perspective, it seems uncontroversial that the listed factors that (alongside other factors) characterise our human form of life have an impact on our practices. These factors are non-normative kinds of facts. So, our normative practices and their rules are shaped by non-normative facts. Now, some might feel the urge to spell out what "shape" and "impact" means in technical terms. They might try to specify the relation that obtains between our normative practices and the non-normative facts of our *Lebensform* in a philosophically rigorous manner. To be in line with Wittgenstein's (meta-)philosophical commitments, this relation needs to be non-reductionistic. So, drawing inspiration from Child (2019; 2020), some

might propose that facts about practices and rules relate to the non-normative facts of our form of life through *supervenience*.[55] A "supervenience" relation obtains if there can be "no difference of one sort without differences of another sort" (Lewis 1986: 14). More technically put, facts about entity A "supervene" on facts about entity B if and only if there can be no difference about the former without a corresponding difference about the latter. There is a strong consensus that reduction requires supervenience, but the latter does not require the former (McLaughlin & Bennett 2021).

If facts about our normative practices and rules supervene on facts about our biology, the environment, and other facts about our human form of life, the following must be the case: if two worlds are alike with regards to the facts about the human form of life, these two worlds must also be alike with respect to the facts about our practices and their rules. However, it seems possible to imagine two worlds which are identical as to their environmental facts but have different practices. Think, for example, about a world that exhibits the same environmental facts as ours but has different styles and, therefore, rules of painting. Advocates of the supervenience idea might reply that our practices and their rules do not supervene on (a) to (d) and all the other kinds of facts that shape our practices individually, but collectively: if two worlds are alike with regards to (a), (b), (c), *and* – rather than or – (d) as well as all the other kinds of facts of our human form of life, these two worlds must also be alike with respect to the facts about our practices and their rules. But is this really the case? Maybe. Yet, to decide this we need to know all the kinds of facts that shape our practices. In other words, we do not know all the facts that would be part of the supervenience base.

Others might think that, instead of supervenience, "metaphysical grounding" might do the job. The term "ground" is often used in reductionist projects; hence, the talk about

[55] Child focuses on the question of whether "semantic facts supervene on non-semantic facts about use" (2019: 229), rather than the question of whether facts about rules and the practices they are embedded in supervene on the facts of the human form of life. Here, I do not make any claims about the question Child addresses. For a critical take on the idea that semantic facts supervene on non-semantic facts about use, see McDowell (1984: 348f.) and Boghossian (2015). For Child's reaction to McDowell's point, see Child (2019: 217–222).

"groundless grounds" above. Yet, in today's metaphysics, "metaphysical grounding" is a technical term that stands for a relation that is weaker than reduction. However, there is little consensus in the emerging literature on metaphysical grounding about the exact nature and structure of this relation: "Despite current enthusiasm for ground, the literature on it is young and very much still in progress. Indeed, there is a lack of consensus on what the central questions are and what methodologies should guide attempts to answer them" (Raven 2020: 12).

For these reasons, I am agnostic about how to characterise the relation that obtains between facts about practices and their rules on the one hand and the facts about our forms of life on the other in technical terms. I do not categorically rule out the possibility that facts about rules can be related to facts of our form of life through some form of supervenience or metaphysical grounding – or through a hitherto unspecified relation. Maybe there are multiple kinds of relations at play.[56] Maybe it is impossible for us to know the nature of the obtaining relation or relations due to our cognitive limits.[57]

Besides that, the above-mentioned facts of our forms of life (a), (b), (c), and (d) are not totally self-enclosed, but, at least partially, interconnected. One might, for example, argue that our conventions about musical tuning are connected to facts about our anatomy as they serve, among other things, the purposes of increasing the playability of the relevant instrument and facilitating the production of musical patterns that are pleasant to the human ear. Furthermore,

[56] Similarly, Williamson who maintains that meaning supervenes in some way on use (1994: 205–209; for a slightly revised version see 1997: 927) does not specify the exact supervenience relation and notes that "[m]eaning may supervene on use in an unsurveyably chaotic way" (Williamson 1994: 209). This may not only pertain to the relation between meaning and use, but also to the relation between human practices and their rules on the one hand and facts of the human form of life on the other.

[57] Hints of this agnosticism can also be found in *Philosophical Investigations*: when touching on the idea of whether "the formation of concepts [or, more generally, of rules] can be explained by facts of nature", Wittgenstein makes clear that this is irrelevant to his philosophical project: "our interest does not fall back upon these possible causes of the formation of concepts [or rules]" (PI: 230). Yet, in line with our agnostic perspective, Wittgenstein does not claim that it is absolutely impossible to reductively explain (the formation of) rules in terms of facts of nature – this matter is simply beyond "our interest". Wittgenstein's agnosticism seems, at least in part, also to be due to his metaphilosophical views. For him, questions about reduction are beyond the remit of philosophy: "we are not doing natural science; nor yet natural history" (PI: 230).

our architectural conventions relate to the affordances of the materials available in the environment: building a Corinthian capital, for example, requires materials with specific petrological features. Apart from that, our desires and volitions relate to facts about our biology and anatomy: if we were immortal creatures, our will to survive would be pointless. I am also agnostic about how to specify the relations that obtain between the various facts of our form of life. Maybe some or even all of them can be characterised in terms of supervenience, metaphysical grounding, or even reduction, maybe not. Maybe the network of relations is so complex and messy that it would be impossible for us humans to make sense of it. Maybe there is a hidden "master fact" (a), (b), (c), and (d) are related to. This might all be the case – but it is most probably beyond our human capacities to find definite answers to these questions. So, we do better to refrain from getting lost in such speculations and, instead, work with what is available and intelligible to us. For, as Wittgenstein has it, "what is hidden[…] is of no interest to us" (PI: §126). From our deflationary, practice-based perspective which prevents us from going down metaphysical rabbit holes, "nothing is hidden" (PI: §435).

A common critique that is often directed against Wittgenstein's anti-reductionism (and anti-reductionist philosophical projects more generally) is that they resist being integrated into a naturalistic worldview. Broadly speaking, naturalism holds that every fact can be explained in terms of natural facts or, by giving it a more Quinean spin[58], the facts of natural science. However, my account does not rule out that it is, at least in principle, possible to "naturalise" facts about rules and practices. We may differentiate between two variants of naturalism which can be labelled "strong naturalism" and "weak naturalism". According to strong naturalism, every fact can be reduced to natural facts or the facts of natural science. According to weak naturalism, every fact is related to natural facts or facts of natural science, but the relation in question is a weaker one than reduction, for example, supervenience or metaphysical

[58] Quine famously contended that "it is within science itself[…] that reality is to be identified and described" (1981: 21).

grounding. As noted above, we are agnostic about whether facts about rules supervene on or are grounded in facts outside of the normative practices they are part of, including natural facts or the facts of natural science. Likewise, we are agnostic about whether facts about rules can be reduced to facts beyond the practices they are embedded in. For it would be absurd to believe that we have reached the endpoint of scientific discovery, that there will be no more scientific breakthroughs in the future. So maybe there are heretofore undiscovered natural facts or scientific principles that enable a naturalistic explanation of rules. Thus, my practice-based, pragmatically agnostic account is, at least in principle, compatible with both strong and weak naturalism. (Yet, it does not presuppose that strong or weak naturalism are true.) However, it is far from obvious that the natural facts and scientific principles that facilitate a convincing naturalistic account of rule-following (provided such facts or principles exist in the first place) can be grasped by the human mind. Consequently, we should draw on what is available and intelligible to us, rather than what is either (hitherto) "undiscovered" or non-existent.

The presented practice-based, deflationary, and pragmatically agnostic take on rule-following is not only, at least in principle, compatible with a naturalistic worldview. It also does justice to the fact that Wittgenstein acknowledges the impact the "natural history" (PI: §§25, 415) of our species has on our normative practices, but is, at the same time, highly critical of any philosophical attempt to reduce the latter to the former (PI: 230). In addition, the pragmatically agnostic thrust of my account of rule-following allows us to avoid attributing metaphysical claims to Wittgenstein (e.g., about the existence or non-existence of certain entities) which would conflict with the anti-metaphysical remarks in his later writings. At the same time, it enables us to avoid reading Wittgenstein as an uncompromising "quietist" whose philosophical endeavours are exclusively negative.[59] Such interpretations are often criticised

[59] Interpreters like, for example, Baker (2004), Horwich (2012), and some of the *New Wittgensteinians*, including Hutchinson and Read (2008), argue that the late Wittgenstein does not engage in *any* positive philosophical project and that his sole aim consists in dispelling confusions and perplexities. Some passages of Wittgenstein's later works may support such an interpretation. But, as noted above, that does not mean that Wittgenstein needs to be read in such a way.

for being self-refuting (Williamson 2013). Independent of whether this objection is justified, the idea that philosophy is a solely negative enterprise would, in my opinion, be disappointing. Based on my take on Wittgenstein, it is possible to pursue positive philosophical projects, but without falling back on metaphysics (at least, on metaphysics of a "revisionary" (Strawson 1959: 9) kind). It is possible if we take our normative practices, and the contexts they are embedded in, as the focal point of our investigation.

Practices are normally shared and socially distributed. (The socio-constructivist account of rule-following with its emphasis on communal agreement implies that rule-following practices are *necessarily* social and shared. However, since the socio-constructivist picture has been rejected, it is enough for us to acknowledge that, at least, most practices are social and shared, without making any claims about necessity.) Moreover, practices which are, in part, shaped by environmental factors provide a frame of reference that does not dissociate creative processes from their environmental context. Hence, by using practices as the frame of reference of our account, it does justice to requirements (3a-b), as outlined in section 3.2.

4. Creativity Surveyed: A New Framework

4.1 A Perspicuous Representation of Creativity: Creativity Spaces

How does the practice-based account of rule-following presented in the previous chapter help us to make sense of creativity? We have seen that the theories or explanations of creativity that have been discussed in the previous two chapters exhibit serious shortcomings: it has been shown that they rest on misguided pictures of rule-following and fail to account for certain examples of creativity. Moreover, it has been shown that "creativity" defies being analysed in terms of a practically feasible and comprehensible analytic definition. As mentioned, due to the limitations of the theories or explanations of language he criticises in his later works, Wittgenstein pronounces: "We must do away with all *explanation*" (PI: §109). (In line with the agnostic remarks above, we do not claim that a theory or explanation of "creativity" is impossible in the strictest sense of the word; yet, if such a theory of explanation is possible, it would most probably be beyond our cognitive grasp.) We can add that we must also do away with all essentialisms – at least, when thinking about creativity in somewhat general terms. But does that mean that we can only speak about creativity in negative terms as a radically "quietist" interpretation would suggest? No, it does not. For, when remarking that we should do away with explanations, Wittgenstein adds: "description alone must take its place" (PI: §109). In brief, to make sense of creativity, we should abandon the attempt to explain or construct a theory of creativity and, instead, adopt a descriptive approach.

Wittgenstein highlights that his descriptive approach proceeds "not by giving new information, but by arranging what we have always known" (PI: §109). Rather than explaining a certain concept in terms of something else (which requires giving new information), the descriptive approach looks at how we use this concept in our everyday lives. As we have seen, the suggested practice-based account of rule-following implies that competent language users, due to their immersion in our linguistic practice, normally know how to correctly use words and concepts that are relevant to them (that are part of their form of life). So, for Wittgenstein,

the "work of the philosopher consists in assembling reminders" (PI: §127) of something we already know, namely of how we use certain words and concepts. In doing so, "[p]hilosophy simply puts everything before us, and neither explains nor deduces anything.—Since everything lies open to view there is nothing to explain" (PI: §126). (For that reason, Wittgenstein states that one "might also give the name 'philosophy' to what is possible *before* all new discoveries and inventions" (PI: §126).)

However, *how* should we arrange what we already know? We assemble reminders of what we know "for a particular purpose" (PI: §127). The purpose of definitions and explanations consists, at least in philosophy, in gaining a better understanding of a certain concept, of its "nature" and the mechanisms associated with it. Since successful analytical definitions and reductionist explanations are not available when it comes to "creativity", we need to find an alternative way of attaining a better understanding of "creativity". For Wittgenstein, a "main source of our failure to understand is that we do not *command a clear view* of the use of our words" (PI: §122). A concept "can only appear in the right light when one has attained greater clarity" (PI: §81) about how it is used. So, despite the absence of an adequate analytical definition or reductive explanation, we can obtain (or, at least, try to obtain) a better understanding of a certain concept by getting a "clear view" of how we use it. However, we are often "lacking in this sort of perspicuity" (PI: §122) that is required for such a clear view, for understanding. Thus, to gain a better understanding of a concept, we need to arrange it in a clear and perspicuous way. That means in such a way that the use or uses of this concept are represented perspicuously. So, to speak with Wittgenstein, we need a "perspicuous representation" (PI §122) or "surveyable overview" of the use(s) of the concept. Wittgenstein emphasises that this concept is central to his philosophical method: "The concept of a perspicuous representation is of fundamental significance for us" (PI §122). It replaces reductionist explanations and analytical definitions, but still allows us to obtain a better understanding of a certain concept: "It earmarks the form of account we give, the way we look

at things" (PI §122). Therefore, this chapter aims to present a perspicuous representation of "creativity" that gives an overview of the different ways in which we use this concept.

For Wittgenstein, it is important that a "perspicuous representation" is *not* a "definition", "explanation", or "theory" as, for him, these terms are linked to essentialism and reductionism. Of course, these terms are sometimes used differently (think, for example, of how the poststructuralists use the term "theory"); yet, for the sake of clarity, I will adopt the way Wittgenstein uses these terms. So, I will abstain from using any of these terms when referring to the surveyable overview that will be delineated in the following. However, as we have seen, Wittgenstein describes "perspicuous representation" as "the form of account we give". In line with that, I will sometimes refer to the perspicuous representation of creativity that will be outlined as an "account" of creativity. Furthermore, I will also use the term "framework" when speaking of this perspicuous overview since this term normally does not bear essentialist or reductionist connotations.

A perspicuous representation allows us to make sense of creativity in a non-essentialist, non-explanatory, non-reductionist way. But how can we obtain such a perspicuous representation? Some Wittgensteinians of a Cavellian stripe suggest that we need to specify "under what circumstances, in what particular cases, do we say …?" (Cavell 1979: 30), with '…' being a placeholder for the concept or term we want to represent perspicuously (see, e.g., Conant 1998: 249f.). This is, of course, not wrong. This strategy helps us to diagnose and, eventually, dissolve conceptual confusions. It helps us to therapeutically treat "philosophical problems [that] arise when language goes on holiday" (PI: §38). Yet, simply listing all particular cases in which we use the term "creativity" (and its cognates) would come close to a Sisyphean task. Moreover, such a list of circumstances and particular cases would not be very illuminating; it would lack perspicuity. Wittgenstein remarks that "[w]e want to establish an order in our knowledge of the use of language: an order with a particular end in view" (PI: §132). So, while the Cavellian strategy might be useful for therapeutic projects, it is of little help for our purpose.

(That does not mean that the perspicuous representation we aim to present cannot be used for therapeutic ends. Yet, it aims to be more than merely a therapeutic tool.)

So, how might we achieve a perspicuous representation suitable for our purposes? In order "to establish an order in our knowledge of the use of language", Wittgenstein stresses that "we shall constantly be giving prominence to distinctions which our ordinary forms of language easily make us overlook" (PI: §132). This is in line with the Cavellian strategy; but Wittgenstein adds another point: a "perspicuous representation produces […] that understanding which consists in 'seeing connexions'" (PI: §122). When it comes to the question of how to make such connections visible, Wittgenstein emphasises the "importance of finding and inventing [interlinks]" (PI: §122).[60] Thus, a perspicuous representation (at least, one that tries to be more than a mere list of circumstances and particular cases) needs to combine two aspects: first, it needs to highlight distinctions we are already aware of (due to our immersion in language), but often overlook when philosophising or theorising about the concept in question. Secondly, we need to do so in a way that allows us to recognise certain connections in how the relevant concept is used. For that, we need to introduce "interlinks". These interlinks should help us to see important connections – as well as important differences.

Since philosophy should only describe and assemble what we already know, Wittgenstein, as we have seen, holds that doing philosophy "is possible *before* all new discoveries and inventions" (PI: §126). However, that does not mean that philosophy, thus understood, must not be "inventive". Indeed, Wittgenstein stresses the importance of "finding and *inventing*" (PI: §122; emphasis added) interlinks. To get a perspicuous overview, we need to "find" or "invent" new concepts or forms of representation that highlight both connections and differences in how the concept in question is being used. Wittgenstein notes that this inventive dimension "may make it look as if we saw it as our task to reform language" (PI:

[60] In the original, Wittgenstein uses the term "*Zwischenglieder*[…]" (1975: §241). Anscombe translates this as "intermediate cases"; yet, "interlink" seems like a more precise translation.

§132). Yet, the later Wittgenstein is highly critical of any attempt to reform ordinary language (which is one reason why he rejects the picture of language he presented in the *Tractatus*). Reforming language is not compatible with the late Wittgenstein's commitment to description, to gaining a perspicuous overview of how we actually use language: "Philosophy may in no way interfere with the actual use of language; it can in the end only describe it. […] It leaves everything as it is" (PI: §124).

However, the conceptual "interlinks" that will be introduced do not aim to reform our actual uses of the concept of "creativity". Quite to the contrary, they aim to render these uses perspicuous – by highlighting both connections and differences in how this term is being used. So, the "interlinks" that will be "found" or "invented" help us gain a perspicuous overview that is in line with the various uses of the concept of "creativity". Accordingly, for Wittgenstein, "improvement[s] in our terminology" or, we may add, inventions of new terminologies, are permissible if they help us "to prevent misunderstandings in practice" (PI: §132) or otherwise enable us to attain perspicuity.

The perspicuous representation that will be presented is not the only possible way of perspicuously representing the complex concept of "creativity". In line with that, Wittgenstein writes: "[w]e want to establish *an* order[…]: an order with a particular end in view; one out of many possible orders; not *the* order" (PI: §132). The concepts and forms of representation (that is, the "interlinks") that will be introduced to render "creativity" perspicuous are "invented". One could also invent different concepts or choose different forms of representation. However, as we will see, the concepts and forms of representation that will be used here help us to pursue our purpose, namely to present an anti-reductionist and non-essentialist account of the concept of "creativity" that is in line with the requirements spelt out in section 3.2. With respect to other purposes, other representations, differently ordered, might be more suitable.

To be in line with requirement (1a), as outlined in section 3.2, an adequate perspicuous representation of "creativity" should not rely on positing necessary and jointly sufficient

conditions. Moreover, to satisfy requirement (1b), it should account for the vague character of "creativity", i.e., the fact that we often disagree on whether something is "creative" and that there are borderline cases of "creativity". We have already touched on the vagueness of "creativity" in sections 2.2 and 3.1. Here are some more examples. Think, for example, of radical avant-garde works, particularly, those that do not manage to break through into the mainstream, that do not find their way into galleries and auction catalogues. People might disagree over whether such pieces are "creative" or merely "original nonsense", to use Kant's (2001: §46) famous phrase. Moreover, although many people would probably reject the idea that a technically perfect replication of Van Gogh's *Café Terrace at Night* which evinces extraordinary painterly skills is a "creative" achievement, not everybody would agree. The question of whether a perfect copy of *Café Terrace at Night* is "creative" might be contentious – however, the question of whether the original *Café Terrace at Night*, painted by Van Gogh in 1888, is a "creative" accomplishment is much less controversial. So, it seems as if some things are more stereotypical (non-controversial) cases of "creativity" than others. Van Gogh's original is, for instance, more stereotypically "creative" than a masterful copy of it (which represents a borderline case). Newton's *Philosophiæ Naturalis Principia Mathematica* is more stereotypically "creative" than a commendable bachelor's thesis in physics. An improvisation by John Coltrane is more stereotypically "creative" than a jazz improvisation performed by a mediocre saxophonist.

What is considered (stereotypically and non-controversially) "creative" is often influenced by one's upbringing, one's belonging to a social group, and one's cultural background. People who are familiar with the history of Western art and music (and their rules) are more likely to consider Malevich's *Black Square* or twelve-tone music "creative" than people who are not. For the latter, it might just be meaningless noise. Social, cultural, and educational differences (that is, differences in intra-human forms of life, as it were) are a source of disagreement about what is "creative". Another source are personal tastes and preferences –

even musicologists might argue about whether John Cage's *4'33* is a "creative" achievement or charlatanry.

For that reason, an adequate perspicuous representation of "creativity" should highlight that some cases of "creativity" are more controversially "creative" than others. Finally, a satisfactory perspicuous representation is perspicuous – and, ideally, more perspicuous than Wittgenstein's *Philosophical Investigations* if one regards it as a collection of perspicuous representations of terms like "language", "meaning", "I", etc. In other words, our perspicuous overview strives to be more than "a number of sketches of landscapes which were made in the course of these long and involved journeyings" (PI: vii). Our perspicuous representation should allow us to see connections and differences – and ideally in an elegant and effective way.

A perspicuous representation that accomplishes this and satisfies the requirements outlined in section 3.2 does justice to the diversity of the phenomena we call "creative". It gives an overview of the different ways in which we use the concept "creativity" (and its cognates). In doing so, it does not interfere with how the concept of "creativity" is actually being used: "It leaves everything as it is" (PI: §124). At the same time, it offers new – or, more adequately, oft-neglected – perspectives on "creativity". A multi-faceted concept like "creativity" that defies essentialist definitions and reductionist explanations may be compared to an ambiguous *Gestalt* image like, for example, the duck-rabbit drawing which is prominently featured in Wittgenstein's *Philosophical Investigations* (PI: 194). The duck-rabbit image can be regarded as the picture of a duck or the picture of a rabbit. Wittgenstein calls this phenomenon "aspect-seeing". When one sees a duck, one discerns the duck-aspect, as it were; when one sees a rabbit, one perceives the rabbit-aspect. At first glance, many people who have not seen the reversible figure before tend to recognise only one aspect – yet, after a while, many beholders also notice the second one. Wittgenstein calls this *aha* moment "change of aspect" (PI: 195f.) or "'dawning' of an aspect" (PI: 194).

However, sometimes, people are incapable of seeing certain aspects. We may call this a symptom of "aspect-negligence".[61] In cases of aspect-negligence, one neglects certain aspects. One only sees the rabbit, for example, but not the duck. Wittgenstein connects his reflections on aspect-seeing to language (PI: 214ff.). Therefore, we can also speak of "meaning-negligence".[62] In cases of "meaning-negligence", one fails to see or neglects certain aspects of the meaning of a concept. So, in a sense, the essentialist definitions and the reductive explanations that have been criticised in sections 2.2, 2.3, and 3.3 are "meaning-negligent". They all fail to account for some kinds of "creativity" and, thus, neglect certain aspects of its meaning.

Our perspicuous representation enables a "change of aspect", the "dawning" of "meaning aspects" that are often neglected. When we recognise an aspect of a *Gestalt* picture we have not noticed before, we "see that it [the picture] has not changed; and yet [we] see it differently" (PI: 193). Similarly, an adequate perspicuous representation "leaves everything as it is" as it does not change how we actually use the term "creativity", but still makes us see it differently. It makes us see meaning aspects we have neglected before. In doing so, it makes us aware of the multifacetedness, the various aspects, of the concept of "creativity". At the same time, it shows how these various aspects hang together. Thus, it discloses unity as well as variety.

However, how to devise such a perspicuous representation that satisfies the criteria outlined above? Even before Wittgenstein there was another luminary who challenged the essentialisms of the philosophical tradition by highlighting the abounding borderline cases and fuzzy phenomena in the world – in his case, in the animal world. The intellectual pioneer in question is none other than Charles Darwin. As Dennett (2017: 138f.) stresses:

[61] Wittgenstein himself does not speak of "aspect-negligence", but of "aspect-blindness" (PI: 213). In most cases, he seems to use this term to refer to the inability to perceive a *change* of aspect, rather than the inability to perceive one of the aspects (Mulhall 2014: 31). However, what we call "aspect-negligence" refers to the latter.
[62] Analogously, Wittgenstein speaks not just of "aspect-blindness", but also of "meaning-blindness" (Wittgenstein 1980b: §189)

> "one of Darwin's most important contributions to thought was his denial of *essentialism*, the ancient philosophical doctrine that claimed that for each type of thing[…] there is an *essence*, a set of necessary and sufficient properties for being that kind of thing. Darwin showed that different species are historically connected by a chain of variations that differed so gradually that there was simply no principled way of drawing a line and saying (for instance) dinosaurs to the left, birds to the right"

To describe all these fuzzy Darwinian phenomena, Godfrey-Smith (2009) suggests a form of representation consisting of three-dimensional arrays, forming a cube, he calls Darwinian Spaces (also see Dennett 2017: 135–149). Before that, Sandra D. Mitchell (2000: 263) developed a similar three-dimensional representational framework to capture the various dimensions of scientific law. My perspicuous representation of creativity is, at least partially, inspired by some aspects of Godfrey-Smith's Darwinian Spaces and Mitchell's multi-dimensional spaces. This does not contradict our critique of Darwinian models or other naturalistic explanations of creativity in chapter 2. For Godfrey-Smith's and Mitchell's frameworks only inspire the *form* of representation underlying our perspicuous overview. This does not oblige us to accept Darwinian or naturalistic theories of creativity. In allusion to Godfrey-Smith's label, I will call my perspicuous representations "Creativity Spaces". Mitchell states: "The world is complex and so must be our scientific representations of it" (2000: 263). Likewise, a satisfactory perspicuous representation of "creativity" should allow for sufficient complexity to capture the concept's multifacetedness – without getting lost in this multifacetedness. What I label "Creativity Spaces" aims to accomplish this.

Before I outline the structure of my Creativity Spaces framework, there are some preliminary remarks to make: in line with the deflationary, practice-based approach to rule-following that has been outlined in the previous chapter, practice serves as the fundamental unit of our investigation. Practice represents the frame of reference within which creativity happens – and, therefore, the frame of reference within which creativity should be examined. Even though it is controversial whether novelty and value are strictly necessary conditions of "creativity", many, probably even most, instances of "creativity" exhibit some sort of novelty or value; and novelties do not pop up *in vacuo*, they need to be situated within a broader context.

Unsurprisingly, our practice-based framework uses our human practices as the frame of reference within which novelty (of whatever kind) occurs. Likewise, on our account, values are embedded in practices.

It seems uncontroversial to say that practices feature certain values. For our practices have significance, serve a certain purpose, within our human form(s) of life. Some moral philosophers even maintain that "[i]t is hard to deny that our evaluative experiences—both what we value, and the ways in which we value things—are profoundly shaped by social practices" (Wallace 2005: 1).[63] Some practices feature opposing values. Respect for human life is, for instance, an important value of the practice of medicine, but not of the practice of warfare. At the same time, some practices share certain values. Respect for human life is, for example, also an important value of other practices such as humanitarian aid.

Our deflationary take on rule-following conceives of practices as imbued with normativity. On this account, as discussed in section 3.3, there cannot be a rule without a practice, nor a practice without rules. However, this raises the question of how to conceive of the difference between the "rule" and the "value" of a practice. Here are some clarifying remarks: what we call the "values" of a practice stand for a certain species of "rules" embedded in practices. The term "rules", in contrast, encompasses all normative standards or recurring meaningful patterns of a practice – including, but not restricted to, "values". So, "values" are "rules" (namely, very important and basic ones in the context of the practice they are part of), whereas not all "rules" are "values". The "values" of a practice can be understood as "foundational" normative standards of a practice that have an impact on the other rules of the respective practice. (But that does not mean that "values" fully determine the other rules of the same practice). "Values" are normally rules that relate to the purpose(s) of the relevant practice. One may, for example, say that intelligibility, efficiency, and – if one agrees with Grice (1975;

[63] Also see Raz (2005).

1989) – cooperation represent fundamental rules – or "values" – of the practice of linguistic communication. The rule for how to construct the past perfect progressive tense, on the other hand, is a rule, but not a value of our linguistic practice.

A good way to characterise what sets "values" apart from other rules is to look at how we tend to react when someone neglects them: if someone is, for example, incapable of correctly forming the past perfect progressive tense, while using language correctly apart from that, we would not say that this person does not speak. On the other hand, if somebody only utters meaningless gibberish – and, thus, neglects the value of intelligibility – many people would deny that what this person is doing qualifies as speaking. As mentioned, the values of a practice usually have an impact on the other rules of the same practice (so that the latter sometimes only make sense in the light of the former). The rule for how to form the past perfect progressive, for instance, allows us to stress the duration of a past event that preceded another event of the past. In doing so, this rule enables us to speak in a more nuanced way about the past – in line with the above-mentioned values of our linguistic practice, that is, intelligibility, efficiency, and cooperation.

Even though the values of a practice shape its face and many of its other rules, that does not mean that I invoke some sort of value essentialism, as it were. For the values of many practices can sometimes change. Consider, for example, the circumstance that, today, twelve-tone music is valued by many people and regarded as part of the practice of music. However, centuries ago, people would not have shared this evaluation; rather, they would have questioned whether such compositions really deserve to be called "music". Even if the values of practices can sometimes be changed, this does not happen very often. In section 3.3, we drew on Wittgenstein's metaphor of a "riverbank" to distinguish between crucial and less crucial rules: in line with this metaphor, the crucial rules of a practice that are hard to change or eradicate have been compared to the "hard rock", while the less important rules that are easier to alter or remove have been compared to the "sand" of the riverbank. The values of a practice are usually

difficult to change (even though it is not impossible). Therefore, they are normally "hard-rock rules", as it were. However, "value" and "hard-rock rules" are not synonyms as there are "hard-rock rules" that are not "values". $E = mc^2$, for instance, is a crucial ("hard-rock") rule of the practice of modern physics that cannot be neglected by any serious physicist. However, it is, obviously, not a value. (More on the values of the practice of science later.) However, in practice, the difference between what is a "value" and what is merely a "rule" is not always clear-cut.

Moreover, some values of certain practices are a matter of discussion and, therefore, not universally accepted. The case of twelve-tone music can again be used to illustrate this point – for, even today, compositions in this style are not valued by everyone who participates in some way in the practice of music (which includes not only performing, but also listening to music). Some practices harbour persistently conflicting values. Compare, for examples, the values that underlie conservative conceptions of art with those that manifest in more progressive ones. Or consider the different values at the core of "analytic" and "continental" philosophy.

Both novelty and value, when viewed within the context of our practices, are gradual phenomena. They come in degrees. Some things are more novel within a given practice than others. And some things are more valuable, that is, more in line with the values of a practice.

When it comes to the ascription of value and novelty in our artistic practices, established "styles" and "genres" play an important role. A "style" or "genre" can be described as a cluster of (often flexible) rules that recur in a specific practice. Impressionism, Cubism, and Suprematism are, for example, styles in the practice of painting; the *Bildungsroman*, the Gothic Novel, and the Sonnet are genres in the practice of literature. Styles and genres also tend to defy being analytically defined in a practically feasible way. For, if one instance of a style or genre lacks one feature which most other entities that exhibit the same style possess, that does not necessarily mean that we do not consider this example a realisation of the respective style or genre, given there are still enough resemblances with other instances of the same style or genre.

Therefore, we should conceive of styles and genres as networks or accumulations of rules that shape their character and generate some degree of resemblance between different objects that belong to a certain style or genre.

The *Philosophical Investigations* briefly touches on "style[s] of painting" and raises the following questions: "is […] our style of painting arbitrary? Can we choose one at pleasure? (The Egyptian, for instance.) Is it a mere question of pleasing and ugly?" (PI: 230). In line with what has been discussed in section 3.4 as well as Wittgenstein's remarks preceding the quoted passage, we cannot reduce painterly styles, as well as other artistic styles and genres, to arbitrary conventions, facts about our desires (e.g., the desire for aesthetic pleasure), or other kinds of facts – even though these kinds of facts have an impact on our artistic practices and, thus, the styles and genres they host. Rather, styles and genres need – just as the practices they are part of – to be understood within the context of the complex form of life they are embedded in. Furthermore, established styles and genres are linked to important values of our artistic practices. (Yet, given that artistic practices sometimes host opposing values, styles or genres need not be in line with *all* the values of the practice they are part of.) The rules of artistic styles and genres play a key role when we evaluate and discuss artistic works. Accordingly, Wittgenstein states that "[i]n aesthetic controversy 'beautiful' & 'more or less beautiful' are hardly ever used; whereas words like 'correct' 'right' are: e.g. that doesn't look quite right yet" (Wittgenstein 2016a: 340). Or as he puts it in a different place: "It is remarkable that in real life, when aesthetic judgements are made, aesthetic adjectives such as 'beautiful', 'fine', etc., play hardly any role at all. […] The words you use are more akin to 'right' and 'correct'" (Wittgenstein 1966: 3).

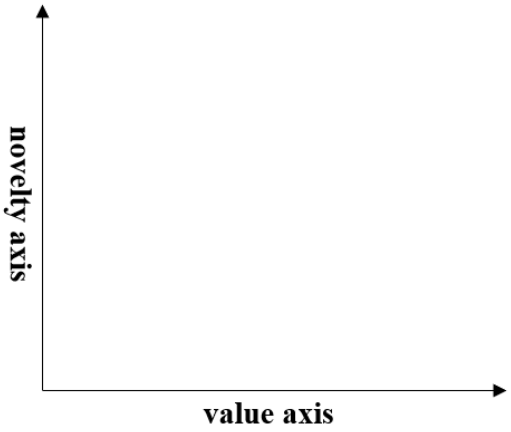
In the sciences, established theories, methods, assumptions, experimental constellations, and instrumentation play a similar role. In the words of Thomas Kuhn: they serve as "paradigms" for what good research looks like. Kuhn famously calls research that is based on the prevalent scientific paradigm "normal science". Occasionally, Kuhn also draws on "rules"

and “standards” to define “normal science”; accordingly, he writes that researchers who “do” normal science are “committed to the same rules and standards for scientific practice” (2012: 11). Lakatos criticised Kuhn’s assumption that scientific practice is dominated by a single paradigm; instead, he speaks of “*competing research programmes (or, if you wish,* ‘*paradigms*’*)*” (2004: 155). It seems dogmatic to state that science as a whole rests on a monopolistic single paradigm or a plurality of partially opposing paradigms. Instead, it seems reasonable to assume that some areas of research are shaped by a single shared paradigm, while others (especially rather new ones) feature rivalrous research programmes and are, thus, shaped by different paradigms. Yet, to bypass these issues and to be coherent with our overall approach, I will characterise scientific work that is considered “correct” (or “normal”) in terms of rules. However, to highlight the rule-based thrust of this characterisation and avoid the paradigm-monopolistic connotations associated to the Kuhnian label “normal science”, I will refer to such research as “standard science”. What I call “standard science” is in line with accepted rules and standards of scientific practice; it is in line with established scientific theories, methods, assumptions, experimental constellations, and instrumentation. This characterisation allows for the idea of opposing scientific approaches and commitments (relating to competing paradigms or research programmes) that are still considered “correct” since a practice can, as mentioned, feature opposing rules. (Nevertheless, even competing research programmes share certain rules, most notably, very basic rules or values of scientific practice; more on that later.)

On this basis, we can introduce the Creativity Space framework. As stated in section 3.2, a satisfactory account needs to account for the synchronic and diachronic dimensions of “creativity”. First, we will focus on creativity in its *synchronic* dimension.

4.2 Synchronic Kinds of Creativity

Let us start by drawing two axes, namely a *Value Axis* and a *Novelty Axis*. Conceiving of value and novelty in terms of axes represents their gradual character. In line with our practice-based approach, both value and novelty are judged against the backdrop of the specific practice the potentially creative entity in question is part of. Let us focus on a random practice that I shall call practice p. Something that tends towards 0 on both axes might be part of practice p – but it would most probably not be considered "creative" (at least, not in a non-ironical manner). To put it crudely, such a thing would be rather boring or useless. On this basis, we can distinguish between three synchronic kinds of creativity.

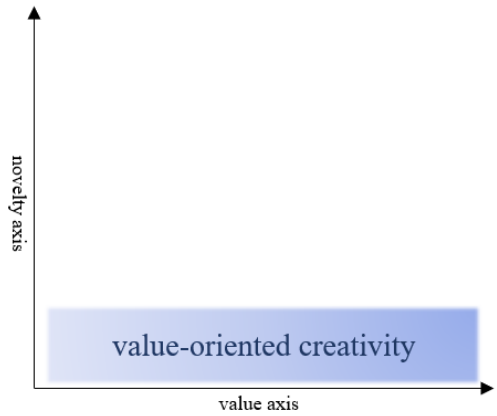


(a) Value-Oriented Creativity

Let us, for a moment, go back to the masterful replica of Van Gogh's *Café Terrace at Night* which represents a borderline case of "creativity". But why is it a borderline case? Why is its "creativity" controversial? Because it is merely a copy and does not embody something uncontentiously novel.[64] But why would some people still ascribe some "creativity" to it? Because it is valuable – it manifests certain values of the practice of art such as aesthetic pleasure and painterly skill. In brief: it represents a borderline case of "creativity" as it is considered valuable, but not new. (Of course, it is not considered as valuable as Van Gogh's original – but, in line with that, the latter is no borderline case, but a rather uncontroversial instance, of "creativity".)

[64] Of course, one might call such a replica "new" insofar as the replica, viewed as an object, did not exist before it was painted (by means of copying). Yet, this will not convince anyone who denies that the *Café Terrace at Night* replica is "creative".

The *Café Terrace at Night* replica represents an extreme case of the first synchronic kind of creativity. We may call it *value-oriented creativity*. Instances of value-oriented creativity are contributions to a practice that are valuable but exhibit only a limited degree of novelty when viewed against the backdrop of the rules of the relevant practice. In other words, value-oriented creativity results from applying the established rules of a practice, without introducing new ones. Even though a borderline case like the masterful replication of an artwork is an (extreme) example of this kind of creativity, there are also less contentious instances of value-oriented creativity. Think, for example, of someone who paints a picture in the style of Impressionism today that is not merely a replication of an existing Impressionist painting. Such a painting is a new object, but we still would not consider it truly "novel" or "original" since it exhibits the rules of an established style of the practice of painting that barely anyone today would regard as "novel" or "original". Similar things can be said if someone writes a novel that strictly emulates the style and genre conventions of a classical *Bildungsroman*, without fully copying an existing *Bildungsroman*.



As mentioned above, Wittgenstein stresses that we often use "words like 'correct' 'right'" and phrases like "that doesn't look quite right yet" or "that won't do, it isn't right" (Wittgenstein 2016a: 340) when evaluating or discussing artworks, that is, instances of artistic creativity (also see Wittgenstein 1978: 3). Instances of value-oriented creativity manifest established rules and values of a given practice. Hence, artworks that are examples of this kind of creativity are commonly judged in terms of "correctness" and "rightness". That does not mean that Wittgenstein holds that terms like "right" or "current" are the only legitimate expressions to use when discussing art. Works of art that manifest different, non-value-oriented kinds of creativity sometimes evoke different reactions. Nevertheless, *many* artworks –

especially those that exhibit value-oriented creativity – are often evaluated in terms of "correctness" and "rightness".

So far, we have only used artistic examples, but in the sciences, we also encounter value-oriented creativity. For example, when scientists use an established explanatory framework, scientific technique, or experimental arrangement – and produce predictable results. More technically put, when scientists apply the established rules of scientific practice and produce expectable results in doing so, we can speak of scientific value-oriented creativity. For that reason, scientific value-oriented creativity is part of "standard science" as it is in line with widespread rules and standards of scientific practice.

However, value-oriented creativity is not restricted to artistic and scientific practices. Many other practices such as chess, cooking, etc. also feature cases of value-oriented creativity. For example, when a chess master defeats her opponent very elegantly and masterfully by only using established strategies.[65] Or when a chef cooks an entrée that has not been "invented" by himself but requires a great amount of skill.

In all these cases, something valuable is brought about: an aesthetically pleasing painting, scientific work that is in line with the values of the practice of science as established scientific methods are being applied, victory in a chess game, a delicious meal. However, established rules were applied to achieve these things. Therefore, the "amount" of novelty they exhibit is limited or, with respect to borderline cases, approaching 0. Instances of value-oriented creativity can feature *some* degree of novelty, but only in a limited sense, especially when compared to examples of the other kinds of creativity that will be discussed later.

Cases of value-oriented creativity often display great mastery of an established (and valued) technique. So, when we use the term "creative" in a value-oriented sense, we often use

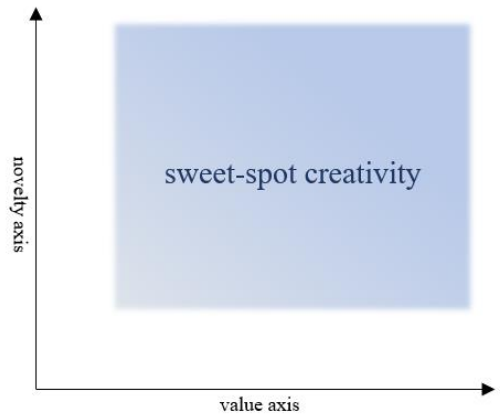
[65] Unlike Koestler, we treat "strategies" as rules of chess as they are the meaningful, recurring patterns in the practice of chess. For the practice of chess is more than merely the "hard" rules that separate legitimate from illegitimate moves.

it as roughly synonymous with adjectives such as “skilful” or “masterful”. Some people might protest and claim that this is not creativity. And certainly not everybody will use “creativity” (and its cognates) in such a way. (And they are free not to use this concept in such a manner.) Nevertheless, the concept of “creativity” is frequently used in such a way. As a corpus-linguistical study conducted by Jordanous and Keller (2016: 18) shows, the term “creativity” is often used in a way that links it to domain competence, skill, and expertise.

(b) Sweet-Spot Creativity

If someone paints a painting in the style of Impressionism in the late 1920s, that might be more novel than doing so today – however, it would not be *very* novel since the style of Impressionism has, at that point, already been around for circa five decades. However, when Claude Monet, Pierre Auguste Renoir, Alfred Sisley, and the other early Impressionists created Impressionist paintings, their pictures embodied something very novel within the practice of painting. And soon they were also considered valuable. So, these early Impressionist paintings are representatives of a second synchronic kind of creativity which we label *sweet-spot creativity*.

Sweet-spot creativity stands for “creative” entities that are, at least to some degree, both valuable and new. In terms of rules: sweet-spot creativity breaks with or challenges the established rules of a practice in a meaningful and valuable way or adds a new set of rules to the practice. (The scale on which rules are broken and the degree to which rule-breaking is appreciated can vary, of course.) Accordingly, the early Impressionists broke with the rules of academic painting that dominated the practice of painting back then. Similarly, while the above-



mentioned *Bildungsroman* represents a case of value-oriented creativity, Goethe's *Wilhelm Meister's Apprenticeship* embodies a case of sweet-spot creativity.

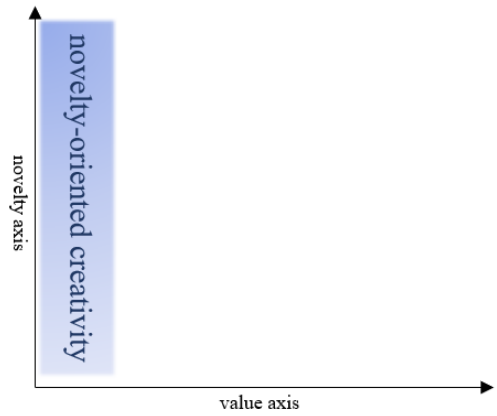
We also encounter sweet-spot creativity in the context of scientific practices. A prime example of scientific sweet-spot creativity is James D. Watson's and Francis Crick's discovery of the double helix structure of DNA. This discovery introduced new rules to the practice of biology and science more broadly. And when Watson and Crick presented their model in their famed *Nature* publication 'Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid' (1953), it immediately proved highly valuable because it was not just in line with the values of scientific practice but was also quickly recognised as a major scientific breakthrough by virtually everyone participating in the practice.

As these examples show, in the arts, the introduction of a new style or genre is a characteristic manifestation of sweet-spot creativity. In the sciences, the development of a new theory or introduction of a novel scientific framework represent typical instances of sweet-spot creativity. But, like creativity of a value-oriented kind, sweet-spot creativity is not restricted to artistic and scientific practices. Non-artistic and non-scientific examples of sweet-spot creativity include, for instance, a highly original and effective chess move that goes beyond traditional strategies in the practice of chess or an entrée that is both a flavourful pleasure and a novelty.

Instances of *sweet-spot creativity* are more likely to be considered stereotypical and uncontentious cases of "creativity" than instances of the other two kinds – unless the "amount" of value or novelty they exhibit is relatively low. Moreover, instances of sweet-spot creativity can be more novel than valuable – and *vice versa*. Our Creativity Spaces framework can capture this.

(c) Novelty-Oriented Creativity

To get an idea of the third synchronic kind of creativity, think, once again, about a radical avant-garde artwork most beholders consider nothing more than "original nonsense" in the Kantian sense, while some deem it more valuable than the majority of people. Such an artwork is an instance of what we call *novelty-oriented creativity*. To put it in more general terms: a "creative" entity that is novel within the context of a given practice, but not considered valuable by most of its participants, is a manifestation of what we will refer to as novelty-oriented creativity. In other words: novelty-oriented creativity breaks with existing rules of a given practice – but in a way that is (by many) not appreciated as valuable. Often, many people cannot even make sense of instances of novelty-oriented creativity. That is why examples of this synchronic kind of creativity are often perceived as more or less "valueless". Accordingly, many instances of novelty-oriented creativity (such as the radical avant-garde artwork just mentioned) are borderline cases of "creativity". There are also numerous examples of scientific novelty-oriented creativity: rejected scientific papers, theories that are deemed too crazy by the majority of participants of scientific practices. The idiosyncratic theories by Lowell, Tesla, and Herschel that have been mentioned above represent extreme examples.



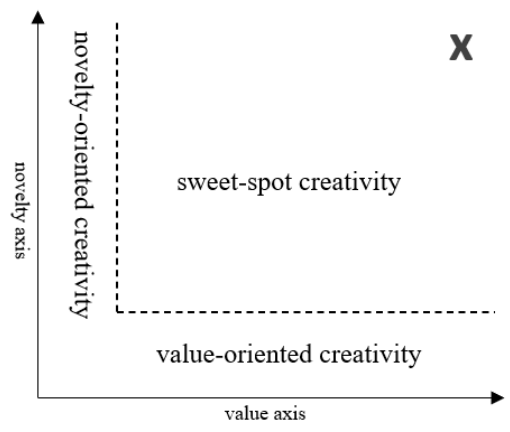
Most instances of novelty-oriented creativity that have been produced in the history of art and science have fallen into oblivion. Their value was too contentious to become canonical. However, not all instances of novelty-oriented creativity have faded into obscurity. We will talk about some of these examples in due course. Like the other two synchronic kinds, novelty-oriented creativity is not only a matter of our artistic and scientific practices. Think, for instance, of a chess player who uses a strategy that is very original, but whose effectiveness is questionable. Or imagine a cook who prepares an entrée that is highly novel but tastes irritating.

Incomprehension and irritation – in extreme cases even shock – are common reactions to novelty-oriented creativity. Another widespread response consists in questioning whether the "creators" of things that exhibit novelty-oriented creativity are "true" representatives of the relevant practice. This response is particularly widespread in our scientific practices: scientists who produce work that is creative in a novelty-oriented sense are often not accepted as "real" scientists by the other participants of the relevant scientific practices.

Creativity Spaces represent what counts as valuable and new within certain practices. Thus, they can only represent value and novelty within a certain practice at certain points of time within a certain geographical and cultural context. What is considered valuable and new in the Paris of 2022 was not necessarily valuable and new in the Paris of 1927 or the Beijing of 2022. Even "within a culture", so to speak, people (who share the same intra-human form of life, as it were) often disagree about what is "creative". As noted above, this is often due to different social backgrounds. So, every Creativity Space has a time and space index and requires us to specify its scope, i.e., the groups it focuses on. As mentioned, Creativity Spaces represent "creativity" synchronically.

(d) Cutting-Edge Creativity

In the picture of our schematic Creativity Space to the left, there is a "x" in the upper right corner. This x stands for achievements that are both extremely novel and extremely valuable within the practice they contribute to. Such creative accomplishments can be referred to as "milestones" in the context of the practices they are part of. This very rare kind of sweet-spot creativity can be called *cutting-edge creativity* (in



the double sense of the word). It radically breaks with existing rules or adds radically new ones in a maximally valuable way.

As we have seen, Wittgenstein observes that we often use words like "correct" and "right" when talking about artworks. As noted above, this particularly applies to value-oriented creativity. Yet, it also applies to some instances of sweet-spot creativity (more on that in the next section). On the other hand, instances of novelty-oriented creativity are often dismissed as they seem "incorrect" or "not right" when judged against the backdrop of established rules. However, in his reflections on art and aesthetics, Wittgenstein admits that we do not always judge artworks solely in terms of "correctness" and "rightness". For some artworks break with, at least, some established aesthetic rules and standards. To exemplify this, Wittgenstein points to Beethoven's symphonic work: "When we talk of a Symphony of Beethoven we don't talk of correctness. Entirely different things enter" (1966: 7f.). For Wittgenstein, artworks that go beyond given aesthetic rules tend to be "the *tremendous* things in Art", the things that "ma[k]e a great impression" (1966: 8). Using our terminology, one can say: while instances of value-oriented creativity tend to trigger comments about their "correctness" and instances of novelty-oriented creativity comments about their "incorrectness", instances of cutting-edge creativity often evoke reactions of amazement, excitement, and astonishment. They appear correct – but in their own way.

What Wittgenstein refers to as "the *tremendous* things in Art" (1966: 8) seems to be roughly equivalent to artistic cutting-edge creativity. So, Beethoven's symphonies can be seen as an instance of this kind of sweet-spot creativity. What are further examples of cutting-edge creativity? Watson's and Crick's discovery of the double helix, for instance. Besides that, it is very likely that, at least in a Western academic setting, figures such as Pablo Picasso, Wolfgang Amadeus Mozart, Louis Armstrong, William Shakespeare, Virginia Woolf, Thomas Alva Edison, Marie Curie, Isaac Newton, or Albert Einstein are mentioned, figures whose works and achievements are widely celebrated as pinnacles of the practices they contributed to. And some

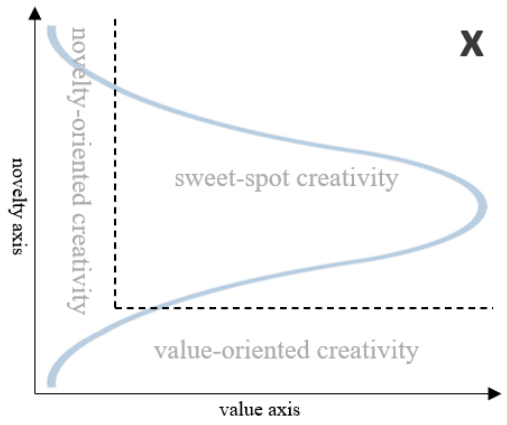
works we associate with these towering figures certainly fall into the category of cutting-edge creativity.

However, things are often more complicated. When first introduced, Picasso's early Cubist paintings or Einstein's theory of relativity, for example, were met with refusal by the majority of recipients. Similarly, "[t]he first exhibition in Munich of the work of Norwegian painter Edvard Munch opened and closed the same day because of the strong negative response from the critics" and some of the "greatest scientific papers have been rejected not just by one, but by several journals before being published" (Sternberg 2003: 109). The recipients and reviewers were, of course, aware that what they saw was radically new – but most of them could not see the value of what they were exposed to because they could not make sense of it. What they saw broke too radically with the established rules of the relevant practices. (Einstein's theory of relativity, for example, "had, at the beginning, far more repercussions in the humanities and arts than genuine testable consequences on physics" (Hacking 2012: xiii).) It took a while until these contributions to their respective practices made sense to the audience and other participants in the relevant practice, were regarded as valuable, and were acclaimed as (extraordinarily) "creative" achievements. They were not immediately recognised as instances of cutting-edge creativity.

As all these examples illustrate, what later becomes cutting-edge creativity often starts as novelty-oriented creativity. To most participants of the relevant practice, they initially appear as "original nonsense" lacking value. In line with that, as a first reaction, they are often denied the status of being "genuine" contributions to the practices they are supposed to be part of. David Hilbert's proof of his finite basis theorem, for example, famously prompted Paul Gordan, as legend has it, to remark: "This is not mathematics; this is theology" (quoted from Corry 2004: 141). After a while, however, Gordan recognised the ingenuity and value of Hilbert's approach, leading him to state: "I have convinced myself that theology also has its merits" ["Ich

habe mich überzeugt, daß auch die Theologie ihre Vorzüge hat" (quoted from Corry 2004: 141, fn 11)].

In the nineteenth century, Wilhelm Wundt (1874) already noticed the circumstance that many of the most impressive examples of (what we now call) "creativity" were initially met with refusal. Wundt became interested in how the arousing potential of a stimulus (which is high when the stimulus is unfamiliar) correlates with its pleasantness. So, in a sense, he was interested in how the novelty of a stimulus relates to its perceived value. Wundt suggested a bell-shaped curve to capture these correlations. I have projected this curve, the so-called *Wundt curve*, into our schematic Creativity Space. According to Wundt, a stimulus is perceived as most pleasant or valuable if it is somewhat new, but not too new.



We make no claims here about whether this always holds true. But it seems uncontroversial that, in many cases, Wundt's observation is accurate. This illustrates that it is extremely difficult to achieve cutting-edge creativity. For, if a stimulus is overly new, there is a strong tendency that it is perceived as unpleasant or valueless. Therefore, an extremely new stimulus is more likely to represent an instance of novelty-oriented creativity than cutting-edge creativity. This seems to, partially, explain why so many historical examples of cutting-edge creativity went through a phase as novelty-oriented creativity before they were recognised as both novel *and* valuable.

To obtain a better understanding of how an instance of novelty-oriented creativity can turn into an instance of cutting-edge creativity, we have to attend to the historical development of such examples of creativity. Since Creativity Spaces are synchronic, i.e., restricted to a certain point in time, they cannot capture the diachronic dimension of creativity. To do justice

to the historical dimension of creativity, we need to differentiate between three further *diachronic* kinds of creativity.

4.3 Diachronic Kinds of Creativity

By introducing the concept of "language-*game*", Wittgenstein uses the metaphor of "game" to reflect on our linguistic practice. In line with that, we will use the notion of "game" as a metaphor for our practices more generally to differentiate between three crucial kinds of *diachronic* creativity.

(a) Game-Playing Creativity

Game-playing creativity leaves existing practices (in their respective historical form) almost entirely unaltered. An example of artistic game-playing creativity would be Impressionists working in the twentieth century: their works are based on a style which is considered valuable, although this style is no longer considered radically novel. More generally, we speak of artistic game-playing creativity when a work of art follows the rules of an established artistic style or genre without altering them. As these examples show, game-playing creativity is normally linked to value-oriented creativity since objects that are creative in this sense do not break with established rules or introduce new ones but rely on the application of already existing rules.

To envision the wealth of artistic, more specifically literary, game-playing creativity, it is worth attending to a study by literary scholar Franco Moretti which shows that most British novel genres between 1740 and 1900 persist for roughly 25 years before they lose their allure and disappear: "forms change once, rapidly, […] and then repeat themselves for two–three decades" (2005: 18). In allusion to Kuhn, Moretti speaks of "normal literature" (Moretti 2005: 18-22). Most works of "normal" or "standard literature" are instances of game-playing creativity as they adopt already established stylistic rules and conventions and, thus, preserve the literary status quo rather than challenging it.

Let us now turn to scientific game-playing creativity. Providing further experimental support for an already well-established scientific theory does not alter the relevant scientific practice but preserves the status quo so that we are dealing with an example of game-playing creativity here. Due to the close connection between value-oriented creativity and game-playing creativity, scientific game-playing creativity abounds in what we call standard science. According to Kuhn, "normal" research revolves around "determination of significant fact [e.g., by quantitatively capturing a phenomenon a theory leaves insufficiently described], matching of facts [experimental data] with theory, and articulation of theory [e.g., by describing the implications and consequences of a theory]" (Kuhn 2012: 34). It seems uncontroversial to regard these activities as important parts of standard science (irrespective of whether this list is exhaustive as Kuhn claims). Most of the time, these standard science activities leave the scientific practice they are part of unchanged, especially if the relevant theory is already well established. Thus, there is an abundance of game-playing creativity in standard science. Nonetheless, despite the close connection between standard science and game-playing creativity, the former is not limited to the latter; more on that shortly.

(b) Game-Changing Creativity

Game-changing creativity changes an existing practice by breaking with some of its rules or introducing a new set of rules. Instances of this kind of creativity may defy established rules and standards, but they set their own rules and standards. Game-changing creativity is normally linked to sweet-spot creativity. This becomes particularly evident in cases of cutting-edge creativity as these are likely to change a given practice.

However, manifestations of sweet-spot creativity can vary in terms of novelty and value. Accordingly, the "scale" or "dimension" of game-changing creativity can vary considerably: some instances of this diachronic kind of creativity change a practice more radically and impactfully than others. To use Wittgenstein's metaphor of the riverbank (which represents a

given practice): while some instances of game-changing creativity only wash away the sand (minor rules), others shatter the hard rock (the crucial rules and values of the respective practice). Some of the most radical examples of game-changing creativity are (or become) instances of cutting-edge creativity such as Picasso's early Cubist paintings, Schönberg's first twelve-tone compositions, or Einstein's theory of relativity.

These achievements dramatically transformed their respective practices. Creativity of this very radical kind is normally linked to achievements that (over time) become cutting-edge creativity. Incomprehension, bewilderment, or shock are common reactions to such radical "game-changers" before they receive acclaim. At first, they are often only meaningful to a small group of people, but gradually they become more popular and accepted. Technically put, they gradually turn from novelty-oriented to game-changing creativity – because they change the rules of the practice they are part of and, thus, the standards according to which they are judged.

However, not all instances of radically game-changing creativity are initially treated with contempt by most participants of the relevant practices, i.e., not all examples of this diachronic kind of creativity start as novelty-oriented creativity. Sometimes, the game-changing impact of creative efforts is overlooked. The revolutionary implications of Planck's work on blackbody radiation in 1900 which later turned out to be a milestone on the way to quantum theory was initially not regarded as ground-breaking, neither by the scientific community, nor by Planck himself (Nickles 2017). It was perceived as a valuable contribution to physics, but not as a radically novel or disruptive one. In brief, before the revolutionary implications – and the cutting-edge creativity – of Planck's work became apparent, it was considered an instance of "standard science" for at least two decades. However, some examples of cutting-edge creativity are immediately recognised as such: when Watson and Crick published their 1953 paper on the double helix, for example, it was promptly considered valuable – and instantly changed the practice of biology and other sciences in a drastic way.

In the arts, extremely impactful instances of game-changing creativity (that concern the "hard rock" of the "riverbank" of their respective practice) often establish or pave the way for a new style, genre, or artistic technique. In the sciences, impactful instances of game-changing creativity often establish or lead to a new theory or a new (sub-)area of research. In his later work, Kuhn distanced himself from the account of scientific revolutions he presented in *The Structure of Scientific Revolutions*, according to which a dominant theory is replaced by a new theory. According to his later account, in contrast, scientific revolutions make a scientific field split into new sub-fields, divisions, and specialities. Accordingly, Kuhn states that the "biological parallel to revolutionary change is not mutation[…] but speciation" (1991: 8). Revolutionary science à la Kuhn can be understood as a manifestation of *radical* scientific game-changing creativity. We are not presenting an account of the "essence" of scientific development or "progress" here. So, there is no need to prioritise "mutation" over "speciation", or *vice versa*. Yet, both "mutation" *and* "speciation" happen in science: new theories are being developed that replace or, at least, refine existing ones *and* new sub-fields come into being. And both things tend to change the relevant scientific practice.

Scientific game-changing creativity is, however, not limited to theoretical work; new experimental designs or methods that prove useful and are widely adopted can also transform a scientific practice. The same is true of observations that do not fit the prevailing theories and models. The latter might not immediately lead to a change in the rules of the relevant scientific practice. Yet, if it cannot be shown that the theory-defying data are defective, they can be seen as evidence against the relevant theory and trigger further research that may, eventually, result in the modification of the theory or the formation of a new one. This can be illustrated by reference to the Michelson–Morley experiment which cast serious doubt on the aether theory and, in doing so, raised a problem that, ultimately, gave rise to the development of special relativity.

It can now be shown that, despite the close ties between standard science and game-playing creativity, the former is, as mentioned above, not restricted to the latter, but can also feature game-changing creativity. In other words: standard science can change a scientific practice. A "standard" experiment whose design follows the rules of the relevant scientific practice can, for example, provide "anomalous" (Kuhn 2012) data that do not fit the established theories and models of the same scientific practice and, in doing so, initiate a line of research that results in an update or replacement of these theories and models. Accordingly, Kuhn emphasises that scientific revolutions are "prepared by the advance of normal science" (2012: 34). It may be an overstatement to claim that this must be the case, but the history of science demonstrates that standard science often sets the scene for scientific revolutions and game-changing creativity more generally.

Unsurprisingly, non-standard science that breaks with some established rules of scientific practice, e.g., by using new experimental designs, by applying new experimental tools, or by proposing a new theory to explain empirical data in a better way, can also change a scientific practice. In other words: non-standard science can bring about game-changing creativity. However, non-standard science does not necessarily change a scientific practice and thus, does not necessarily produce game-changing creativity. If novel non-standard theories, experimental designs, or research methods do not prevail (for example, due to the circumstance that they remain unintelligible, i.e., instances of novelty-oriented creativity), it is unlikely that they have a lasting impact on the relevant scientific practice.

Practices can be changed in different ways. Most of the above-mentioned examples of game-changing creativity are instances which *radically* changed the practices they contribute to. They changed some of the crucial rules or even some of the values – that is, parts of the "hard rock" – of the practice. However, there are also less radical forms of game-changing creativity. It is impossible to list all the ways in which practices can be transformed, but it is possible to differentiate between several ways in which practices can be less radically changed.

In other words: it is possible to differentiate between different types of non-radical game-changing creativity. However, the following overview does not strive to be exhaustive – it only aims to present some kinds of game-changing creativity. Moreover, it concentrates on artistic and scientific practices. That, however, does not mean that the kinds of game-changing creativity that will be distinguished cannot be found in other practices. Lastly, it is also worth mentioning that the various types that will be delineated in the following are not mutually exclusive; some creative achievements exhibit various of these types.

(i) Problematising Rules

As highlighted above, some instances of standard science like, famously, the Michelson–Morley experiment point to certain problems or anomalies of standard science, i.e., of the established rules within scientific practice. Such scientific work "problematises" the status quo. In doing so, it often paves the way for further research that brings about more dramatic changes (some of which will be outlined below). One might argue that the arts are, as it were, less "problem-centric" than the sciences. However, in our artistic practices, it also happens sometimes that certain rules or styles become "problematic". Work in aesthetics, literary criticism, or manifestos often problematise given aesthetic rules and styles. Think, for instance, of Adolf Loos's (2012) criticism of ornamentation which gave crucial impetus to modern architecture.

(ii) Establishing New Rules

Artistic examples of this type of game-changing creativity are early *Bildungsroman* novels that were not among the very first examples of this literary genre, the works of the early Impressionists that did not "invent" this artistic style, or the musical work of Ray Charles, Sam Cooke, and Otis Redding. All these artworks were not the first of their kind, i.e., they did not originate the styles they exhibit or genres they represent. Likewise, Joyce's *Ulysses* was not the

first display of the stream-of-consciousness technique. All these artistic examples did not introduce the rules of the respective styles and techniques they manifest but helped "popularise" them. They helped to establish styles and techniques that were, in a sense, already present in the respective artistic practice, but unknown to many people (even to those familiar with the relevant practice). They modified the practices they are part of, but they did so by drawing on rules which the relevant practice already "contained" in some form.

Scientific work that helps to establish a contentious theory also exemplifies this type of game-changing creativity. A spectacular example is Eddington's highly influential experiment from 1919 which provided observational evidence for the then-controversial theory of general relativity which, as a result, triumphed over the Newtonian status quo. If Eddington and others had not delivered experimental data in favour of Einstein's *Relativitätstheorie*, it might have remained an instance of novelty-oriented creativity that is too controversial to fundamentally change the practice of physics. The history of Bayesian statistics (a branch of statistics which treats probability as a degree of belief) provides another example. Thomas Bayes himself never published the work that sets the foundation for Bayesian statistics. After Bayes's death, Richard Price discovered Bayes's unpublished essay "An Essay Towards Solving a Problem in the Doctrine of Chances", recongised its revolutionary potential, and arranged its publication. Price edited and even amended Bayes's paper to some extent. In a word: Price did not invent Bayesian statistics, but his work played a crucial role in establishing it as a branch of statistics. (These examples – and instances of the present type of game-changing creativity more generally – showcase the socially distributed character of many creative processes, as highlighted by requirement (3a) in section 3.2.)

(iii) Tweaking the Rules

This type of game-changing creativity can be exemplified with reference to the following Impressionist paintings: *L'Absinthe* by Degas and *Le Jugement de Pâris* by Renoir. These two

artworks transgress some rules of traditional Impressionism, as it were: orthodox Impressionist paintings have no contour line; in Degas's painting, however, the absinthe glass, the carafe, the tables, the dress of the melancholically glancing lady are all framed by contour lines. Moreover, prototypical Impressionist paintings show everyday scenes like busy boulevards, promenades, parks, and gardens. Renoir's painting, in contrast, displays a mythical scene. Nonetheless, these paintings by Degas and Renoir still breathe the spirit of Impressionism. So, even if these two paintings strain established rules of the painterly style they adopt, they do not transform this style in a more fundamental way. Accordingly, people generally accept these two artworks as Impressionist pictures. More generally, the present type of game-changing creativity stretches or tweaks certain rules of a given style, but without inducing any major mutations of the relevant style, let alone the practice the style is embedded in (as crucial rules and values of the relevant practice remain untouched). Another example of this game-changing kind of creativity are new (i.e., non-established, unconventional) metaphors: they flout existing linguistic rules and add something to the respective language – but the language itself does not undergo a fundamental change.

The same type of game-changing creativity can also be encountered in the sciences. Evolutionary theory as first developed by Darwin contained several mistakes, e.g., the assumption that species diversity increases exponentially over time which is mainly due to the influence that Malthus's (2008) work on population growth had on Darwin (Hsü 1986: 532). It later turned out that this assumption was erroneous. According to Sepkoski (et al., 1981), the pattern of Phanerozoic species diversification is, for instance, more logarithmic than exponential. Correcting Darwin's error helped to refine evolutionary theory, but it did not change this theory and its rules in a fundamental way. In a sense, the rules that were changed by correcting this defect were not crucial to evolutionary theory, but only peripheral.[66]

[66] This differentiation is similar to Lakatos's distinction between the "hard core" of a research programme and its "protective belt" (Lakatos 2004: 133-138).

(iv) Merging Rules

As shown in section 2.3, combinatorial theories argue that creativity results from combining elements from different domains. We have criticised these theories for absolutising combinatorial processes. Yet, even if combinatorial processes are not the only way to make creativity happen, the latter still often results from the former. The current type of game-changing creativity is based on combinatorial processes. There are two sub-types depending on what is being combined or merged.

(α) *Merging Rules within Practices*

Instances of the first "combinatorial" sub-type merge rules within a given practice. In the arts, this often takes the form of combining the rules of different styles or genres that are part of the same artistic practice. Think, for example, of the paintings by the late Lovis Corinth which amalgamate stylistic features of Impressionism and Expressionism, the compositions of Alban Berg that complement the Schönbergian twelve-tone technique with elements of Mahlerian romanticism, or William Blake's *Marriage of Heaven and Hell* which combines various genres, including prose, poetry, aphorisms, and engravings. All these artworks blend rules of different styles or genres. However, the rules of the styles and genres that are mixed were already present in the relevant practice. Corinth, for example, invented neither Impressionism nor Expressionism, but was one of the first and most talented mergers of both painterly styles. In other words: synchronically viewed, the amount of novelty exhibited by instances of this kind of game-changing creativity is often limited. Accordingly, this combinatorial sub-type often does not lead to a radical change of the relevant artistic practice.

The current combinatorial sub-type can also be encountered in the sciences. For example, when concepts, methods, procedures, or theories that stem from different sub-disciplines of the same scientific discipline are combined. In that case, the rules of different sub-practices that belong to the same scientific practice are merged. This can be illustrated with

reference to the technique of "renormalization" that was developed in the 1940s by Julian Schwinger, Freeman Dyson, Sin-Itiro Tomonaga, Richard Feynman, and others to solve fundamental problems in quantum field theory ('t Hooft 2016: 1f.). Simply put, this technique made it possible to calculate finite answers for physical quantities for which other procedures produced infinite results. It did not take long until renormalization was also used in other areas of physics. Physicists like Murray Gell-Mann and Kenneth G. Wilson combined renormalization, for example, with ideas from particle physics and statistical physics (Kadanoff 2013: 174f.). Another example is the application of category theory in various branches of mathematics. Category theory was invented in the 1940s by Samuel Eilenberg and Saunders Mac Lane as a general theory of mathematical structures which was first used in algebraic topology and abstract algebra. Ever since, its areas of application have diversified. Alexander Grothendieck, for example, applied category theory to algebraic geometry and William Lawvere combined it with logic and detected a range of interesting connections (Awodey 2006: 1f.).

(β) *Merging Rules between Practices*

While the examples mentioned in the previous section mix different styles within the same practice, there are also cases of combinatorial creativity that merge the rules of *different* practices. In the arts, the latter sub-type sometimes manifests in the creation of hybrid artforms such as Wagner's music drama, sound film, happenings, or digital art. Yet, there are also more subtle forms: consider, for example, Béla Bartók's use of certain mathematical patterns (or rules) such as the Fibonacci sequence in his musical composition (Lendvai 1979; Maurice 2004: 88ff.). Or think about the application of Cubist features to sculpture after the inception of Cubism in painting. Or consider the adoption of rules of scientific writing by nineteenth-century novelists such as Émile Zola (1894). In the first example, rules of the practice of mathematics are transferred to the practice of music; in the second example, rules of the practice of painting

are transferred to the practice of sculpture; in the third example, rules of the practice of scientific writing are transferred to the practice of literature.

In the sciences, the current combinatorial sub-type sometimes manifests in the formation of a new sub-discipline within a given scientific practice by combining rules of the latter with rules of another (scientific) practice. This can be illustrated by pointing to the invention of quantum computing. The birth of (the core idea behind) quantum computing is often dated back to a 1981 lecture by Richard Feynman who suggested that the most promising approach to simulating nature is by devising a computer that follows quantum-mechanical rules (Brooks 2019: 20; Feynman 1982). Another example is the formation of statistical mechanics by physicists such as Ludwig Boltzmann and James Clerk Maxwell in the late nineteenth century through the combination of classical thermodynamics which is part of the practice of physics with statistical methods, i.e., methods that stem from the practice of mathematics.

The different kinds of game-changing creativity are not necessarily mutually exclusive. In fact, some game-changing creative efforts feature more than one of the types that have been outlined. The special theory of relativity, for example, not only broke with many rules of the practice of physics and introduced new rules – it also combined or merged various (more or less established) rules of the physics of the day, including the rule to treat light speed as a constant and rules that follow from the work of Lorentz and Poincaré.

(c) Game-Inventing Creativity

Game-inventing creativity manifests itself in the creation of an entirely new practice. Examples of this kind of creativity are the invention of painting circa 40,000 years ago, the invention of Western philosophy by the Pre-Socratics (or an ingenious someone or group whose name or names are now forgotten), the Scientific Revolution that led to the emancipation of science

from philosophy, the invention of cinema, or the invention of computing by Ada Lovelace and others.

The distinction between game-changing and game-inventing creativity is not always clear-cut – not least because game-inventing creativity always happens within a wider context (a form of life) and, thus, often draws on aspects of existing practices that play an important role in that context. Nonetheless, in cases of game-inventing creativity, there is a widespread acknowledgement that what is being established represents a new *kind* of practice that is distinct from existing practices. Once they were recognised as valuable novelties, Cubist artworks were (and are still) regarded as art, twelve-tone compositions as music, and the theory of relativity as physics. In fact, those who are critical of, say, Schönberg's work are usually critical precisely because they deny that it represents a contribution to "music" – which implies that even harsh critics of twelve-note music recognise that it is supposed to be a contribution to the practice of "music". In other words: instances of game-changing creativity are judged and thus understood in terms of existing practices and their rules. Instances of game-inventing creativity, in contrast, tend to raise the question of *what* it is that one is doing in the first place. Of course, in the face of this question, it is a common reaction to draw on established practices and their rules in order to make sense of what one is confronted with. Yet, even if there may be overlaps, instances of game-inventing creativity resist being merged into existing practices. That does not mean that instances of game-changing creativity are beyond criticism. However, in contrast to examples of game-changing creativity, the criticisms they raise tend to revolve around the question of whether the *practice* that is being established is really useful and relevant – not within the context of a specific given practice, but more generally within our way of life.

Instances of game-inventing creativity tend to evoke the coinage of a new name to describe what one is doing. But so do many instances of game-changing creativity. As mentioned, game-inventing creativity establishes a new practice, while game-changing creativity changes an existing one. So, to distinguish between game-inventing and game-

changing creativity it is often helpful to ask the question: does it make sense to say that I "participate in" what has been brought about by a certain instance of creativity. If the answer is yes, game-inventing creativity tends to be at play. For example, it makes sense to say that I "participate in" physics, philosophy, photography, computer science, or chess. Yet, it sounds odd to say that I "participate in" Cubism, twelve-tone music, or relativity theory. Of course, I can say that I "participate in" quantum physics, microbiology, analytic philosophy, concept art, or still life photography. However, the first lexical element of these terms makes clear that the whole term refers to a sub-discipline that is part of a larger discipline (or practice) that is denoted by the second lexical element.

Game-inventing creativity often, but not always, takes the form of the invention of a new technology or technique. The practices of photography, cinema, music recording, and writing, for instance, were made possible by the invention of the daguerreotype, motion picture cameras, the phonograph or gramophone, and the alphabet. Moreover, new technologies played an important role in the Scientific Revolution, most famously, the telescope which was invented in 1608 and was used by Galileo and others to provide observational evidence against the Ptolemaic model of the cosmos (Tyson 2007: 76) – which set an example of how to (scientifically) study the universe. Of course, new technologies can also facilitate other kinds of diachronic creativity. The nineteenth-century invention of the metal paint tube, for example, played a key role in the formation of Impressionism – an example of game-changing creativity – as it allowed its protagonists to paint *en plein air*. Nevertheless, highly disruptive and transformative technological innovations (like the alphabet, daguerreotype, or computer) tend to lead to entirely new practices à la game-inventing creativity.

It has been mentioned that new practices that are brought about by game-inventing creativity are often intertwined with other already existing practices. Besides that, many practices comprise other practices, sub-practices, as it were. The practice of science, for example, features the practices of physics, chemistry, biology, physiology, neuroscience, to

name but a few. These practices, in turn, involve further practices, e.g., experimental practices. Likewise, the practice of art includes the practices of painting, sculpture, literature, music, and many more. These sub-practices also include further practices; the practice of painting, for instance, features various painterly practices or techniques. This shows that practices can be viewed at different "levels of granularity", so to speak. If one focuses on certain experimental practices in the practice of neuroscience or specific painterly practices relevant to a specific style of abstract painting, the level of granularity is higher than when one looks at the practice of science or art in general.

Game-inventing creativity on a relatively high level of granularity often represents game-changing creativity on a lower level of granularity. Accordingly, the *invention* of a new experimental practice or method in the practice of microbiology, for example, can change the practice of microbiology – maybe even the practice of biology or science more comprehensively. Likewise, the *invention* of a new painterly practice or technique – such as Janet Sobel's and Pollock's invention of the drip technique or Max Ernst's invention of the grattage and frottage techniques – can change the practice of painting and related practices. (The invention of such sub-practices, as it were, *can*, but need not, change the practices they are part of because they can remain instances of novelty-oriented creativity.)

As should have become clear by now, in the context of this investigation, we concentrate on creativity on a low level of granularity. Accordingly, when we have used the term "practice" so far, we were speaking of academic disciplines, artforms, games, etc. – rather than sub-practices (that is, specialised methods, techniques, or strategies) within these practices. The concepts that have been introduced in this chapter can also be used to analyse creativity on a higher level of granularity.

We have already seen that, by using human practices as the crucial frame of reference, our account is in line with requirements (3a-b), as delineated in section 3.2. In the meantime, it

should have become apparent that it also satisfies the other requirements. As shown, the Creativity Spaces framework does not rely on positing necessary and jointly sufficient conditions for "creativity" and allows for the existence of borderline cases. Thus, it does justice to requirements (1a, b). Furthermore, our account presents synchronic as well as diachronic kinds of creativity – and, thus, does not conceive of "creativity" in an ahistorical manner, as demanded by requirement (2a). In addition, the notion of game-changing creativity accounts for the circumstance that some instances of creativity change existing frames of reference (i.e., existing practices), as suggested by (2b). In line with (2c), the concept of game-inventing creativity accounts for the creative emergence of new frames of reference (practices) that defy being classified as transformed versions of pre-existing frames of reference (practices). Finally, in accordance with (2d), our framework accounts for the circumstance that transforming an existing frame of reference (practice) or inventing a new one are not the only ways of being creative by introducing the notion of game-playing creativity. Our framework also acknowledges that creativity sometimes results from applying established techniques by classifying this way of being creative as a way of displaying game-playing creativity or (from a synchronic perspective) value-oriented creativity. Hence, the presented framework also lives up to requirements (2a-d).

The boundaries between the different synchronic and diachronic kinds of creativity that have been presented are not always clear-cut. The emergence of Pointillism in the late nineteenth century exhibits aspects of tweaking-the-rules creativity and the more far-reaching style-inventing kind of creativity. For Pointillism clearly evolved from Impressionism, but is still somewhat distinct from it. Accordingly, there are obvious similarities and differences between both styles. But it is not a problem that there are not always distinct boundaries: "I strive, *not for exactness*, but for perspicuity" (Wittgenstein 1980b: §895).

4.4 Artistic vs Scientific Creativity

In the previous section, we have discussed several artistic and scientific examples of creativity. This, however, should not hide the fact that there is more to be said about the differences that obtain, various similarities aside, between artistic and scientific creativity. The framework that has been presented in the previous sections does not conceal these differences but allows us to elaborate on them, as this section will briefly illustrate. To accentuate one crucial difference between artistic and scientific creativity, it is helpful to highlight two points of criticism that are occasionally raised against modern art and science with respect to their respective innovative capacities.

As far as today's artistic practices are concerned, there is a widespread impression that art is prone to (dis)plays of self-reflexivity that are incomprehensible if not irrelevant to people outside the artworld. This sentiment finds, for example, expression in the German saying "Ist das Kunst oder kann das weg?" ("is this art or can it be discarded?") which took hold after an artwork by Joseph Beuys had been cleaned up by a custodian who did not recognise it as art (Weibel 2019: 127). Baudrillard goes even so far to claim that the nature of contemporary art consists in "asserting nullity, insignificance, meaninglessness" (2005: 27). At the same time, there is a widespread sense that, in times of highly specialised and detailed scientific research, revolutionary scientific breakthroughs become increasingly rare. In line with that, based on a large-scale analysis of scientific papers from six decades, a recent *Nature* article argued that scientific work is "increasingly less likely to break with the past in ways that push science […] in new directions" (Park et al. 2023: 138). Simonton even suggests that "[s]cientific genius is extinct" (2013: 602).

These views may represent simplifications or even caricatures, but they point to certain tendencies within today's artistic and scientific practices. Using our vocabulary, they complain about the lack of cutting-edge and radically game-changing creativity in the arts and the sciences. However, the situation in both practices is quite different: while current scientific

practices gravitate towards value-oriented and game-playing creativity, contemporary artistic practices are, at least from the outside, often perceived as hotbeds of novelty-oriented (but non-game-changing) creativity. (The latter seems to pertain especially to, but is not limited to, visual artforms.) This difference is, in large part, due to different core values that guide these practices. As mentioned, the "values" of a practice represent its foundational rules that relate to the purpose(s) of the relevant practice and have an impact on its less fundamental rules.

Naively viewed, it may seem as if "truth" is the core value of science. Accordingly, Nietzsche maintains that scientists are animated by the "will to truth" (2007: 112) and Popper considers scientific practice the "persistent and recklessly critical quest for truth" (2002b: 281). However, "truth" represents a difficult and philosophically burdened concept. So, there is a strong tendency in the philosophy of science to bypass this notion and focus on more tangible "epistemic values" or "canons of inference" (Levi 1960: 356).[67] Kuhn (1978a: 321ff.), for example, invokes, among other things, "consistency" as a key value of science. According to this value, a scientific theory needs to be consistent not just internally, but also with other established scientific theories. The latter aspect promotes clinging to, rather than deviating from, the *status quo*. For when choosing between two theories, all other things being equal, the one which is more closely aligned or consistent with accepted theories ought to be favoured. Quine and Ullian point in a similar direction by presenting "conservatism" (1978: 66ff.) as a crucial scientific value (or virtue, as they prefer to say): the lower the degree to which a hypothesis or theory conflicts with existing scientific hypotheses, beliefs, assumptions, or theories, the better. Quine and Ullian add "modesty" as another important scientific value according to which slight and familiar hypotheses and theories are preferrable to more extravagant and obscure ones (1978: 68f.).

[67] In today's philosophy of science, it is uncontroversial to claim that scientific practice is guided by epistemic values like the ones we will concentrate on here. The more controversial question concerns the role of *social* and *ethical* values in scientific inference-making and theory-building. This issue cannot be discussed here; for an overview, see Douglas (2016).

All these values (which partially overlap) have in common that they prioritise the known over the unknown, the familiar over the unfamiliar, the established over the unestablished. In brief: they encourage a conservative attitude. Popper (2002a: 420) and Lakatos (2004: 177) even speak of a "dogmatic attitude" that permeates scientific practice. The space of possible novelties and innovations in science is further limited by another value which Kuhn calls "accuracy" (1978a: 321ff.) and which demands that a scientific theory be in line with empirical data and the results of experiments. All these values promote value-oriented and game-playing creativity as well as mild forms of sweet-spot and game-changing creativity. Analogously, they make it difficult to achieve more ambitious forms of game-changing and sweet-spot creativity (including cutting-edge creativity) in the sciences.

Identifying the values of contemporary art proves more difficult. At least since the emergence of what Karl Rosenkranz calls the *Aesthetic of Ugliness* (2015), art is no longer committed to classical and (Neo-)Platonist values such as beauty, perfection, timelessness, or (aesthetic) truth. Especially since the beginning of modernism, art allows for a wide variety of experiences and ideas. In fact, in contrast to pre-modern aesthetics, new and surprising experiences and ideas are particularly valued in the context of modern art. In other words, modernity established novelty as a key value of art. Or as Gianni Vattimo puts it: modernity, especially modern art, centres around the "value of the new" (2014: 116). This new value of the new was prepared by the emergence of Paradigm 2 thinking (as described in section 2.1): in contrast to Paradigm 1, Paradigm 2 renders it possible for humans to produce genuine novelties. And with the change of aesthetic values that goes hand in hand with modernity, artistic newness is not only possible, but highly valued. The great amount of value attached to novelty in modern art manifests itself in the artistic examples of cutting-edge creativity and radically game-changing creativity that have been discussed above.

There is also cutting-edge and radically game-changing creativity in the sciences. Yet, as Kuhn figures out, "innovation itself need not be a prime value for scientists, and innovation

for its own sake can be condemned" (1978b: 350). As opposed to this, many "artists do seek new things to express and new ways to express them" and they "make innovation a primary value" (Kuhn 1978b: 350). Kuhn even suggests that "the radically different value placed upon innovation for innovation's sake by scientists and artists" (1978b: 350) may represent the deepest difference between both practices.

Artistic novelty can take various forms. It can, for example, materialise in new artistic techniques, new "sujets", the use of new materials and media. It can also manifest in more idiosyncratic ways, for example, in the "transfiguration of the commonplace" (Danto 1981) that characterises certain branches of modern art, most strikingly, pop art; for such an artistic "transfiguration" presents *new* perspectives on both ordinary objects and the "nature" of art. Following Vattimo who proclaims that postmodernity initiated a "dissolution of the value of the new" (2014: 117), it might be tempting to argue that the value of the new is vanishing in contemporary art. Yet, this claim is not justified. For, despite the rhetoric of anti-innovation, postmodern art is not just reproductive or imitative. It might – in theory – contest inherited distinctions (including the distinction between the old and the new), but – in practice – it relies on innovative modes of production or presentation, often including genre bending, style mixing, intermedia explorations, and self-reflexivity.

However, the quest for novelty in contemporary art – especially in combination with the "transfigurative" and "postmodern" (i.e., genre-bending, style-mixing, intermedial, and self-reflexive, etc.) strategies that permeate today's artistic practices – often leads to the production of works whose meaning, let alone value, is – at least, from an outside perspective – not clearly recognisable. So, in practice, the value of the new often leads to the production of novelty at the expense of value (or, more precisely, value that is widely shared). In brief, the value of the new often promotes novelty without value (i.e., easily identifiable value). Technically speaking, in practice, this value often fosters novelty-oriented rather than cutting-edge creativity.

For that reason, in general, creativity tends to take different forms in today's artistic and scientific practices. That, however, does not mean that there is no novelty-oriented creativity in the current sciences (think, for example, about string theory), nor that there is no value-oriented creativity in contemporary art (think, for example, about the persistence of certain literary genres or architectural styles). Nonetheless, the tendencies outlined above highlight different challenges today's artistic and scientific practices are facing when it comes to achieving cutting-edge creativity and radically game-changing creativity: while it tends to be difficult to make radically novel contributions to the sciences (especially more "mature" ones), it is often difficult to produce art that is perceived as meaningful and valuable outside the relevant artistic practice.

The creative tendencies within contemporary artistic and scientific practices that have been sketched above are not set in stone. If our artistic and scientific practices featured different values, it is likely that different creative tendencies would obtain. As mentioned, the values of an established practice can be challenged or even altered. New values are normally introduced to a practice through radical game-changing creativity. Thus, the current creative status quo of the arts and sciences can be changed. The values of a practice normally relate to the wider context the practice is embedded in. As shown, the value of the new that permeates contemporary art would not have been possible, say, in medieval times.

As announced, we have mainly concentrated on artistic and scientific creativity here. However, our framework can also be used to analyse creativity in other practices. Besides art and science, the practices of computing and engineering represent other hotbeds of creativity. Especially AI research and its latest (game-changing) creative achievements attract a lot of attention at the moment. Some AI systems even generate seemingly creative outputs. This raises the question if or to what extent AI-powered systems can be creative.

5. Computing Machinery and Creativity

Today, we are confronted with algorithms that "write" novels or "compose" music in the style of Bach, and witness how AI-generated paintings find their way into auction catalogues and museums. In the face of these and other breath-taking achievements of state-of-the-art AI, more and more people (particularly journalists) are not only convinced that machines can be truly creative, but increasingly concerned that computers might, in the not-too-distant future, outperform humans (that is us) when it comes to creative power. However, a number of theorists reject the hype and assert that "creativity" cannot legitimately be ascribed to machines. In a word: there is a great amount of disagreement and controversy when it comes to the question of whether and to what extent machines can be creative.

Many Wittgensteinians seem to be particularly inclined to hold the view that "creativity" cannot be legitimately attributed to machines. Although Wittgenstein does not explicitly say anything about whether machines can be creative, there are some scattered remarks in his later works about whether machines can think. In the *Blue Book*, Wittgenstein writes:

> 'Is it possible for a machine to think?' ... And the trouble which is expressed in this question is not really that we don't yet know a machine which could do the job .... The trouble is rather that the sentence, 'A machine thinks (perceives, wishes)': seems somehow nonsensical. It is as though we had asked 'Has the number 3 a colour?' (1964: 47)

Here, Wittgenstein writes that it *seems* nonsensical to say that "a machine thinks" – that it is like stating that "the number 3 is blue". In *Philosophical Investigations*, Wittgenstein brings up the same question again: "Could a machine think?" (PI: 359) – but this time his answer is even more straightforward: "a machine surely cannot think" (PI: 360).

For that reason, it is tempting to think that Wittgenstein thinks that it is impossible for machines to think since it is nonsense to claim that they could. And if that is what one thinks, why should one think that machines can be creative? Wittgensteinians that adopt a "nonsense policeman" reading (as described in section 3.4) may believe that a person that calls a machine "creative" – just like someone who calls a machine "intelligent" – misuses language by

violating “grammatical” rules (i.e., rules governing how we ordinarily use a word). In other words: the “grammar” of the word “creative” tells us that this word can be applied to humans (Paradigm 2), maybe to gods (if one has not fully left behind Paradigm 1), but not to machines. In that case, speaking of “creative machines” is nonsense.

As shown in section 3.4, advocates of the nonsense policeman reading maintain that “grammatical” rules capture how we ordinarily use words. Yet, our ordinary uses of words are often more diverse and irregular than such alleged nonsense policemen assume. This also holds true for our use of the word “creative”: some people (and not just non-Wittgensteinian philosophers) use the word “creative” to describe the performance of some machines, while others would not do so. So, in the face of the fact that some people call machines and their outputs “creative”, while others do not, attempts to reject the view that machines can be creative on purely conceptual or grammatical grounds turn out to be biased and misguided. They are no longer viable today. People who are convinced that machines can be creative call machines, particularly AI systems, “creative” in a non-ironical and non-metaphorical way. Thus, *today*, appeals to rules of ordinary language use fail to unmask talk about “creative machines” as nonsense.

When dealing with Wittgenstein’s dismissive remarks on machine intelligence, one has to consider that these were made more than 70 years ago. Accordingly, Wittgenstein comments on his statement that machines cannot think as follows: “We only say of a human being and *what is like one* that it thinks. […] Look at the word ‘to think’ as a tool” (PI: 360; emphasis added). Wittgenstein compares the word “to think” with a tool. A tool can often be used to cope with various *similar* tasks. And, sometimes, an old tool can be used to deal with a new task (which somehow resembles the tasks it has so far been used for). Accordingly, the meaning of a term can change in the course of time – given the right circumstances, it might, for instance, be extended to new cases. Since the language-games we play are embedded in a wider context,

such extensions of the scope of concepts are often due to transformations in the world we live in. As Nyíri describes it:

> [I]f the environment changes in significant ways, our language-games can lose their point [...]. Let us assume that computers at some stage will become able to accomplish, in important areas and on a significant level, feats that would require intelligence if accomplished by people; and let us, further, assume that [due to these technological advances] human attitudes and behaviour with respect to computers will at the same time undergo some relevant, appropriate modifications. One expects that under such conditions a change in the use — a shift in the meaning — of certain psychological terms would occur; that there would arise a tendency to ascribe to computers a kind of thinking (1989: 386).

So, in some circumstances, it may be possible to ascribe a predicate P to an object x if there are enough similarities between x and the things P is ordinarily attributed to. Wittgenstein declares with regard to the ascription of mental or perceptual states: "only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious" (PI: §281). For Wittgenstein, we only ascribe mental or perceptual states to beings that resemble humans. So, concerning the question when it makes sense to say that something is "thinking" or "intelligent", it might be concluded that "if something is sufficiently 'like' a human being it is reasonable to attribute to it the property of thinking" (Penco 2012: 189). The same seems to be true of the ascription of "creativity".

What "sufficiently like a human being" means is, of course, contentious. Those people who (unironically and non-metaphorically) call AI "creative" believe that these systems sufficiently resemble us in important respects. They might point to artworks generated by an AI system and say that, if a human being had produced it, we would not hesitate to call it "creative". Sceptics of computational creativity, on the other side, might object that, though the products or behaviour of some machines might appear to be creative, this "creativity" is not *real* creativity as paradigmatically displayed by humans. They might accept that some people call some machines "creative", i.e., that there is, in some language-games, a rule for calling machines "creative". Yet, they would say that this is a bad or misleading rule for it unites things under one label that do not belong together. For that reason, in their view, it is better to refrain from using the term "creative" to describe machines and their performance.

Several arguments have been developed to vindicate the view that machines are *not* sufficiently like humans to be truly creative, i.e., creative in the sense in which humans are creative. In the following, I will engage critically with the most influential of these arguments. "Critically" is to be taken in a Kantian way here: "Kantian critique is content neither with the role of prosecutor nor that of defendant, fulfilling instead the offices of a judge—specifically, a civil rather than a criminal judge. It imposes no sanctions, but adjudicates a legal title" (Höffe 2001: 17). In the following, several prosecutors will get a chance to bring charges against the idea of machine creativity. These points, however, will be challenged. Finally, after evaluating the strength of the arguments against artificial creativity, it will be judged whether (some) AI systems deserve the "legal title" of being "creative", that is, creative in the human-centric sense of Paradigm 2 thinking.

The following objections to computational creativity contend that AI systems are, for certain philosophical reasons, not able to be genuinely creative. Most of the objections that are (explicitly or implicitly) put forward to challenge the notion of machine creativity fall into two types: (a) the first kind of objection highlights the dependency of the machine's seemingly "creative" performance on human influence or interference – more generally: on human factors. Ada Lovelace was arguably the first sceptic of AI creativity *avant la lettre* to argue along those lines. (b) The second type of criticism consists in highlighting important differences or dissimilarities between machines and humans: it is asserted that humans can do certain things machines cannot, or have specific qualities which machines do not possess, which are necessarily required in order to be creative. The latter argumentative strategy seems to appeal to some Wittgensteinians. In the following, different manifestations of these two kinds of objections will be scrutinised.

Both argumentative strategies to discredit computational creativity proceed by arguing that humans can be truly creative, while machines can, at most, only appear to be creative. In line with Paradigm 2, they treat human creativity as paradigmatic. It will be examined whether

the following arguments against artificial creativity are convincing against the backdrop of the human-centric paradigm. So, to determine if machines really cannot be creative, human creativity will serve, to borrow another Wittgensteinian phrase, as an "object of comparison" or "measuring-rod" (PI: §131). If there are unilateral dependencies between machine creativity and human creativity (or, at least, human factors) or if machines lack certain qualities that are necessary for being creative in a human sense, it seems justified to conclude that machines cannot be creative. If not, it seems justified to say the opposite, namely that machines can be creative.

Let us start with the first kind of objection according to which machine creativity crucially depends on human factors. As noted above, this type of critique can be traced back to none other than Ada Lovelace.

5.1 The Lovelace Objections

Commenting on the Analytical Engine which was conceived by Charles Babbage in 1837, Ada Lovelace who was arguably the first programmer in history stated: "The Analytical Engine has no pretensions whatever to *originate* anything" (quoted from Hartree 1949: 70; also see Turing 1950: 450). In other words, the Analytical Engine cannot be creative. To make this point, Lovelace declares: "It [the Analytical Engine] can do whatever *we know how to order it* to perform" (quoted from Hartree 1949: 70). Turing called this "Lady Lovelace's Objection" (1950: 450): a "machine can only do what we tell it to do" (1950: 454). (While Turing treats the sceptical position he attributes to Lovelace as a rejection of the idea of machine *intelligence*, it is, more specifically, a critique of machine *creativity*.) The *Lovelace Objection* implies that whenever a machine seems to behave in a creative way or produce a creative object, this creativity is actually to be attributed to its creator or the programmer who wrote the programme that runs the machine.

Lovelace's remark might produce the reply that she is talking about a machine of the first half of the nineteenth century – a machine that quintessentially differs from the computers we use today. Yet, this assumption is, as Turing (1950: 450) shows, wrong: although Babbage never completed the Analytical Engine, it is now regarded as the first conception of a mechanical computer. At its core, the Analytical Engine is a general-purpose computer. More technically speaking, it is "Turing complete", which means that it can, in principle, simulate any Turing machine. In other words, such a machine is "computationally universal". As Turing states: "The Analytical Engine was a universal digital computer" (1950: 450) – and as such, if it possessed enough storage capacity and processing power, it could simulate any other Turing machine, including modern Turing machines such as the computers we have today. (For "[a]ll programmable computers in use today are in essence Turing machines" (Dennett 2004: 295).) Therefore, (independently of whether Lovelace herself refers exclusively to Babbage's invention or makes a more general point) the Lovelace Objection, if construed in this more fundamental way, also affects modern information-processing machines.

(a) Lovelace Objection 1.0

According to the *Lovelace Objection*, machines cannot be creative (or intelligent) since they cannot, as Lovelace puts it, "*originate* anything". Yet, what does that mean? It is often taken to mean that, as Turing phrases it, "a machine can 'never do anything really new'" (1950: 450). Let us call this interpretation of Lovelace's complaint the Lovelace Objection 1.0. Yet, already in the 1950s, this view proved to be unwarranted. To point to one example: in 1956 (only six years after Turing published his paper on machine intelligence), Allen Newell, Herbert A. Simon, and Cliff Shaw wrote the Logic Theorist, one of the first AI programmes in history. This pioneering system was not only able to prove most of the logical theorems in the second chapter of Alfred North Whitehead and Bertrand Russell's *Principia Mathematica* – it even delivered a significantly more elegant proof to one of them (McCorduck 1979: 161–170). In

other words, the Logic Theorist was able to "*invent* logical proofs" (Bostrom 2014: 7; emphasis added). It originated a proof which was new and superior to the one presented by logicians of the calibre of Whitehead and Russell. Its creators did not tell the Logic Theorist to come up with this new and more elegant proof – it was also new to them. Sceptics of AI computational creativity might feel tempted to respond by calling for stricter criteria for something to count as "really new". But this reply would be unfair: if human beings achieved the things the mentioned AI system accomplished, most people would not hesitate to call them "creative". Consequently, the Lovelace Objection 1.0 turns out to be untenable.

(b) Lovelace Objections 1.1 and 2.1

As mentioned above, Lovelace also critically notes that the Analytical Engine can only do "whatever *we know how to order it* to perform". Lovelace does not say that a machine like the Analytical Engine cannot originate anything because it cannot do what its creators can or know how to do. Such a claim could easily be rejected: one of the most obvious counterexamples is IBM's Deep Blue, the first chess computer to defeat a reigning world champion. None of its creators would stand a chance in a chess match against Garry Kasparov, one of the greatest chess virtuosos of all time. Nevertheless, Deep Blue managed to do exactly that. Yet, Lovelace states that machines cannot do something we do not know how to *order*. The Logic Theorist can prove logical theorems and Deep Blue managed to defeat Kasparov. If Lovelace is right, these systems can only do these things because their creators know how to "order" these things – that is, because their creators knew how to state and formalise the rules involved when doing logical proofs and playing chess. So, machines like the Logic Theorist and Deep Blue only managed to do these impressive and seemingly inventive things since their creators managed to capture the rules that govern logical deduction and chess in a computer programme. Let us call this criticism the Lovelace Objection 1.1: even though machines can generate "creative" outputs or behave "creatively", they themselves are not "creative" since their performance

requires humans who program them, i.e., provide them with the relevant instructions to perform certain tasks. For that reason, advocates of this objection might demand that the programmers and engineers at IBM who created Deep Blue, rather than the machine itself, ought to be praised as "creative" for the chess moves that eventually led to the victory over Kasparov.

The Lovelace Objection 1.1, however, only affects what Haugeland (1985: 112) called "Good Old-Fashioned Artificial Intelligence" (often abbreviated as "GOFAI"), i.e., traditional, symbolic, logic-based AI systems which rely on programmed instructions. Yet, over the last few decades, the AI community has gradually turned from logic-based to statistics-based approaches. This led to the Machine Learning revolution in AI. The crucial idea behind Machine Learning (ML) is to make the machine "learn the algorithms" that are required to accomplish a task "automatically from data, replacing programmers with learning programs" (Alpaydin 2016: x). Due to the awe-inspiring advances in ML and data science, there has been a shift in AI from programming a machine to do something towards letting it learn itself what to do. ML "gives computers the ability to learn to perform a task from data without being explicitly programmed" (Sejnowski 2018: 282). So, it is not necessary anymore that programmers know and formalise all the rules their programmes are operating under as today's computers can, thanks to ML methods, extract rules from data. Such ML programmes "learn" from examples. Consequently, the programmer has often no idea what is happening "inside" the machine, what rules it operates on. AI researchers often compare such ML systems with black boxes: although we know the inputs and outputs of such an algorithm, we have no clue how it arrives at its decisions. Hence, the Lovelace Objection 1.1 proves to be out of place when it comes to ML systems: such machines can do things we do not know how to order them to perform, as it might be put in allusion to Lovelace's famous remark.

However, the Lovelace Objection can easily be reformulated so as to apply to ML-powered systems. Let us call this the Lovelace Objection 2.1 (the "2" will stand for "seemingly ML-proof"): machines, even if they produce "creative" outputs, can themselves not be "truly"

creative since they are dependent on humans inasmuch as it is humans who feed them with data, connect them to an energy source, etc. Humans cannot be fully removed from the creative process.

However, the Lovelace Objections 1.1 and 2.1 neglect the circumstance that humans often rely on other humans in their creative efforts (as highlighted by requirement (3a) in section 3.2). Furthermore, humans are often dependent on different technologies and tools when they are creative. Without a brush, Velázquez would not have been able to paint *Las Meninas* (and he does not hide it); without a guitar, Jimi Hendrix would not have been able to play his solos; without a camera, Cindy Sherman would not have been able to take her self-portraits, exploring social identity; without a movie camera, Spike Lee would not have been able to make *Do the Right Thing*. This shows that many practices are scaffolded and enabled by technologies. In other words: technologies play a crucial part in our human form of life. Since technologies are integral to many of our practices, Ortega y Gasset describes humans as "ontological centaur[s]" (146: 154). Similarly, Andy Clark calls human beings "natural-born cyborgs" due to "their ability to enter into deep and complex relationships with nonbiological constructs, props, and aids" (2003: 3, 5). There may be creative efforts like telling an innovative joke that do not rely on technologies. However, practically all artistic and scientific practices are scaffolded by technologies. Thus, one applies double standards if one argues that machines cannot be creative because they rely on humans to do so; for, at least in the arts and sciences, human creative efforts also rely on technologies.

(c) Lovelace Objection 2.2

Some of Lovelace's heirs may still not be convinced and invoke another objection. In the era of ML, as we have seen, "data starts to drive the operation; it is not the programmers anymore but the data itself that defines what to do next" (Alpaydin 2016: 11). What might be called the Lovelace Objection 2.2 takes this into account and points to the fact that the data which are

used to train allegedly "creative" ML systems were "made" by humans: ML-driven music-generators learn on music (or musical data, as it were), composed by humans such as Bach or Debussy. Likewise, ML networks with pretensions to painterly artisanship were trained on paintings (or painterly data, as it were), made by human artists such as Monet or Picasso. Lovelace Objection 2.2 takes this to show that ML-based machines cannot be creative.

However, human composers, painters, etc. are normally not creative in a social vacuum, detached from all previous creative achievements of their practice: "We are constantly absorbing the ideas of others. […] Accumulation—building upon what you and others have accomplished—is part and parcel of creative activity" (Miller 2019: 13). Even extraordinarily creative humans are, usually, influenced by previous accomplishments and existing styles and formats in their respective field. As far as value-oriented and game-playing creativity are concerned, this is obvious since humans who are creative in this sense apply established rules of existing styles, which they themselves did not "invent": humans who compose in the style of Bach (without being Bach) or paint in the style of Cubism (without being Picasso or Braque) have usually been exposed to the music of Bach or Cubist paintings and went through some training to master the relevant musical or painterly rules. Even people who are creative in a game-changing way do not work in a sphere unconnected to existing styles and rules. Picasso, for example, before he transformed the practice of painting, thoroughly studied the styles of contemporaries like Cézanne (Brassaï 1999: 107; Loran 2006: 7) and of old masters such as El Greco, Velázquez, and Goya (Brown 1996; Rosenblum 1996; Caws 2005: 31f.). In order to be able to depart from existing styles in a meaningful and valuable way, it tends to be helpful to be familiar with what one is departing from. To break existing rules in such a way that, at least, some listeners and beholders can relate to one's composition or painting normally requires one to be aware of existing styles and their rules.

As a result, the Lovelace Objection 2.2 fails to deny ML systems the ability to be creative. For it is not only machines of this type that are exposed to and learn on examples of

human creativity before they produce (seemingly) creative outputs – the same normally also applies to humans. So, like Lovelace Objections 1.1 and 2.1, Lovelace Objection 2.2 neglects the distributed, scaffolded, and technologically mediated character of most instances of human creativity. Thus, all the variants of the Lovelace Objection we have discussed fail to convincingly demonstrate that machines cannot be creative.

5.2 Dissimilarity Objections: Incompleteness, Consciousness, and Free Will

(a) Incompleteness Objection

Roger Penrose contends that it is impossible for computers to fully simulate human intelligence. To make that point, he refers to the Gödel's first *incompleteness theorem* and the unsolvability of the *halting problem*[68] that show that there are (mathematical) questions a human mathematician would be able to answer, while a Turing machine would not be able to do so (Penrose 1994; 1999; also see Lucas 1961). Critics of computational creativity might feel tempted to argue that this objection not only undermines the possibility of artificial intelligence, but also of machine creativity.

However, Penrose's argument can be refuted by pointing to the fact that many human beings cannot solve all the mathematical problems professional human mathematicians can solve. Nevertheless, we would not say that it is impossible for those people to be intelligent or creative. As if he anticipated Penrose's objection, Turing (the very mathematician who had shown that the *halting problem* is algorithmically unsolvable) writes: "I would say that fair play must be given to the machine. Instead of it sometimes giving no answer we could arrange that

[68] The *halting problem* is a problem in computability and complexity theory. Broadly speaking, it asks whether, given a computer program and an input, the program will halt at some point or continue to run indefinitely. In his seminal 1937 paper 'On Computable Numbers, with an Application to the Entscheidungsproblem', Turing proved that the halting problem is algorithmically undecidable. (By doing so, Turing almost accidentally invented the computer).

it gives occasional wrong answers. But the human mathematician would also make blunders" (1986: 124).

(b) Consciousness Objection

Critics of AI creativity often argue that machines cannot be genuinely creative since they lack certain qualities which are necessary for being truly creative. Unsurprisingly, one of the most prominent candidates for such a quality is phenomenal consciousness or the capacity to have conscious experience (see, e.g., O'Hear 1995). AIs might display what is sometimes called functional consciousness, i.e., architectural structures that resemble those that seem to underly phenomenal consciousness in humans (Schneider 2019: 49). Yet, even if machines exhibit functional consciousness (which some theorists deny the status of consciousness proper), that does not necessarily make it the case that they also have phenomenal consciousness.[69]

However, declaring phenomenal consciousness necessary for the ability to be creative is also problematic. First, despite enormous efforts in the philosophy of mind, there is still little clarity about what phenomenal consciousness is and how it works (Van Gulick 2022). Secondly, it raises solipsistic concerns. In his AI paper, Turing criticises the view that machines cannot be intelligent because they are not phenomenally conscious: "[a]ccording to the most extreme form of this view the only way by which one could be sure that a machine thinks is to be the machine and to feel oneself thinking. […] Likewise, according to this view the only way to know that a *man* thinks is to be that particular man" (Turing 1950: 446). Although Turing is talking about intelligence here, his argument also pertains to creativity. For, since we do not have direct, first-person access to the phenomenal consciousness of other people, one would need to be, say, Frida Kahlo to assess if she was creative. Thus, the view that being creative requires phenomenal consciousness rests on a "solipsist point of view" (Turing 1950: 446)

[69] The same pertains to what Block (1995) calls "access-consciousness", i.e., the availability of the information carried by an inner state to the wider system; on access-consciousness in AI, see Tzafestas (2022: 159f.).

which nobody would like to adopt.[70] (In section 5.3, another argument against the view that being creative requires phenomenal consciousness will be presented.)

(c) Free Will Objection

Another capacity that is sometimes invoked to deny machines creativity is free will. A variant of the view that creativity presupposes free will can be found in Popper (1972: 222).[71] However, like the consciousness discussion, the free will debate is a philosophical minefield: "After all these centuries of writing about free will, it does not seem to me that we have made very much progress" (Searle 2007: 37). We are still groping in the dark as to whether humans, that is, we ourselves, have free will. To bypass this problem, some philosophers shift the focus from metaphysical proofs or rebuttals of the existence of free will to the introspective experience of free will (Lehrer 1960; Ginet 1990; O'Connor 1995; Strawson 2004; also see Bayne 2017). However, the attempt to deny machines the capacity to be creative on the basis that they lack the introspective experience of free will proves also problematic. As we do not have direct, first-person access to other people's introspective experience of free will, this attempt also falls prey to solipsism. To find out if a specific person can be creative, one would, again, need to be this very person. So, appeals to free will or the experience thereof fail to undermine the idea of machine creativity.[72]

The philosophy of the late Wittgenstein with its anti-mentalistic thrust also casts serious doubt on the idea that being creative presupposes phenomenal consciousness or (the experience of) free will. To advert the threat of solipsism, he emphasises that the ascription of allegedly "inner process[es or phenomena] stand[…] in need of outward criteria" (PI: §580).

[70] Pointing to the Surrealist methods of artistic production such as automatic drawing and painting, Boden (2014: 234f.) also argues that consciousness is not strictly necessary for being creative.
[71] For further expressions of this view, see Kane (1998: 81).
[72] Here, we focus on the question of whether creative beings need to *have* free will (or an introspective experience thereof), rather than the question of whether creative beings need to behave like they have free will. Later, we will touch on issues closely related to the latter question.

Nevertheless, several critics of AI still hold that Wittgenstein's later philosophy undermines the idea of machine creativity. So, in the next two sections, we focus on two supposedly Wittgensteinian dissimilarity objections.

5.3 Rule-Following Objection

In chapter 4, different types of creativity have been differentiated. These kinds of creativity have been characterised in terms of how we use and relate to existing practices and their rules: game-playing creativity is the result of applying existing rules (of existing styles and practices); game-changing creativity transforms a given practice, for example, by altering or breaking existing rules and establishing new ones; game-inventing creativity leads to the inception of a new practice. Applying a rule, altering or breaking a rule, and inventing a new rule-guided practice – it all presupposes the capacity to follow rules. So, to find out whether and to what extent these different kinds of creativity are attainable by machines, it is necessary to examine whether computers can follow rules. At the same time, if machines are incapable of following rules, these kinds of creativity – and creativity in general – seem out of reach for them.

In chapter 3, we have already focused on issues surrounding rule-following. More specifically, we have mainly grappled with the constitutive or ontological question of what constitutes correctness when it comes to applying a rule. To a lesser degree though, we have also touched on the epistemological question of how rule-followers know what a rule requires them to do in order to follow it correctly. Here, we concentrate on a third question Wittgenstein also touches on in his later writings, namely the question of who or what qualifies as a rule-follower. This third question might be called the "attributional question" of rule-following.

By referring to Wittgenstein's remarks on rules, several critics of computational creativity such as Peter Hacker (2019) and Stuart Shanker (2002) deny that machines, including AI systems, can be genuine rule-followers. On this basis, another dissimilarity objection to computational creativity – one might call it the *Rule-Following Objection* – can be advanced,

which can be formulated along the following lines: "Machines cannot follow rules; hence, they cannot be creative".

Before we can examine whether this criticism is justified, we have to deal with the following question: when do we, in Wittgenstein's view, say that somebody or something is following a rule? Wittgenstein makes the point that solely behaving in accordance with a rule does not count as following a rule (PI: §232). Accordingly, he differentiates between rule-following and mere behavioural regularity. But what is the difference between following a rule and merely behaving in accordance with a rule? Think of a group of French native speakers who communicate in French to each other. We would say that these people follow linguistic rules. Yet, if a German tourist who does not speak French picks up some French phrases when on holiday in Paris and repeats these phrases, not knowing what they mean, we would not say that this person follows the (semantic) rules of the French language, but only behaves in accordance with them. Imagine the German tourist combines two of the phrases he picked up (without knowing what they mean) and that the Parisians who hear him say that combination think that it exhibits poetic beauty, "creativity". However, since the tourist cannot follow the rules of the French language, we would not say that he is "creative" (at least, not in a non-metaphorical way), although what comes out of his mouth seems to be so. Neither do we regard clouds as rule-followers. Accordingly, even if we see a face or another object in the clouds, we would not refer to the clouds as "creative".[73] Rather, it is us humans who are creative when we see faces in the sky.

The question we are now confronted with is: are machines that produce allegedly "creative" outputs or display allegedly "creative" behaviour just creative in this cloud-like manner, namely not at all? To tackle this question, one has to examine whether or to what extent

[73] One might object that these "faces in the sky" are not "made" by the clouds themselves, but by the wind. However, today it would be equally strange to say (in a non-metaphorical way) that the wind is "creative". For the member of a tribe with an animistic religion, it might make sense to say something like that – but in most contexts, it would not.

machines can follow rules. For machines only seem to be able to achieve human-like creativity if they can follow rules. According to Hacker, however, machines cannot follow, but merely be in accord with, a rule: "A machine can execute operations that accord with a rule, provided all the causal links built into it function as designed and assuming that the design ensures the generation of a regularity in accord with a chosen rule or rules. But for something to constitute following a rule, the mere production of a regularity in accord with a rule is not sufficient" (2019: 106).

To scrutinize whether machines really cannot follow rules, one first has to examine what sets rule-followers apart from beings that can merely behave in accordance with a rule. As we have seen, Wittgenstein convincingly rejects the view that the correct application of a rule can be explained in terms of mental facts. Accordingly, Wittgenstein repudiates the view that genuine rule-following necessarily involves some kind of conscious experience or any other sort of mental state. To stress that, Wittgenstein invokes a thought experiment (PI: §160): he pictures a scenario in which a person who can read (i.e., who is capable of following the rules that are required to do so) is asked to read a text which she does not know. Yet, due to the effects of some peculiar poison, the person, while she is reading the text, has the deceptive sensation that what she is doing is not reading, but reciting something she has learnt by heart. An illiterate child who recites a text it has learned by heart does not follow the rules of reading, but simply behaves in accordance with them. For the protagonist of Wittgenstein's thought experiment, however, it would not be possible to recite the text from memory since she has never been exposed to it before and, consequently, has no memories of it. So, she reads the text and is only under the impression that she is not doing so. Hence, it is possible to follow a rule without having a conscious experience or sensation tied to it.[74]

[74] A similar argument can be found in Proudfoot (2004a: 364f.; 2004b: 286).

Furthermore, the ability to follow a rule does not necessarily presuppose some kind of reflective understanding, propositional knowledge, or interpretation of the relevant rule: in daily life, we often apply rules (think, for example, of syntactical rules), without having any deeper understanding or knowledge of them – not all of us are linguists. Consequently, explicit understanding or knowledge of rules is neither necessary, nor sufficient, for the capacity to follow them.

For Wittgenstein, in order to follow a rule, it is not necessary to have knowledge about it or certain mental states accompanying the application of it. So Wittgenstein does not restrict rule-following to humans: in *Remarks on the Foundations of Mathematics*, for instance, he presents a scenario, revolving around two chimpanzees; one of the chimpanzees "scratched the figure |— —| in the earth and thereupon the other the series |— —||— —|"(Wittgenstein 1978: 345). Does that qualify as rule-following? Wittgenstein writes:

> If […] there were observed, e.g., the phenomenon of a kind of instruction, of showing how and of imitation, of lucky and misfiring attempts, of reward and punishment and the like; if at length the one who had been so trained put figures which he had never seen before one after another in sequence as in the first example, then we should probably say that the one chimpanzee was writing rules down, and the other following them (Wittgenstein 1978: 345).

So, for Wittgenstein, non-human animals can follow rules. And, since Wittgenstein plausibly shows that the ability to follow rules does not rely on mental states, it seems legitimate to assume that even without a mental-states-producing, fleshy brain one can qualify as a rule-follower.

So again: when does somebody or something follow a rule instead of merely behaving in accordance with one? As we have seen in chapter 3, Wittgenstein shows that the correct application of a rule is not fixed by mental, neurophysiological, or consensus-based facts. Rather, Wittgenstein emphasises that rule-following is a practice. Accordingly, as demonstrated in greater detail in chapter 3, rules are neither more basic than the practice they are embedded in, nor are practices more fundamental than their rules. So, a rule-follower can competently participate in a practice, while somebody or something that merely behaves in accordance with

a rule cannot do so. Accordingly, we would not say that the German tourist follows the rules of the French language since he cannot properly partake in the practice of speaking French. At the same time, we are more inclined to accept that the two chimpanzees in Wittgenstein's thought experiment follow rules for they seem to participate in a shared practice: practices are guided by rules, have normative standards – and one of the chimpanzees teaches the other about these rules. By using words like "instruction", "showing how", "reward", and "punishment", Wittgenstein emphasises that one of the chimpanzees introduces the other to certain normative standards, to a practice and its rules. A cloud, in contrast, does not engage with a shared practice and its normative standards.

Since the ability to follow rules is related to one's ability to participate in a practice, the right question one has to ask in order to figure out if machines can follow rules is: can machines participate in practices? And, since we are dealing with the question of whether machines can be creative in the sense humans are creative, we have to ask more precisely: can machines participate in our human practices?

Some critics of artificial creativity with a background in ML might deny that (at least, as far as ML-powered systems are concerned). They may contend that, even if ML-based systems may be capable of following rules in some sense, we often cannot know if they follow the same rules humans follow. And if it is unclear whether ML systems follow the same rules we follow, how can we be sure that they can participate in our human practices? To raise this concern, they might invoke an oft-quoted anecdote about a neural network that was, as legend has it, supposed to learn to automatically detect tanks, but actually learnt to detect clouds. This was due to the circumstance that all the tank-depicting pictures the neural net was trained on by using supervised learning methods were taken on a cloudy day, while the non-tank-depicting pictures were taken on a sunny day (Yudkowsky 2008). This might be an extreme and obviously problematic example. However, due to ML-specific issues such as "overfitting" and "dataset bias", we often cannot fully control what an ML-based system learns. Since training datasets

are always limited in size, they are almost always (at least to some extent) biased. Furthermore, whenever there is noise in the data, "overfitting" occurs. "Overfitting" means that "a model is too specific to the features of the training data" as it "learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data" (Vermeulen 2020: 65). As it is often not possible to free datasets from noise, it is often not possible to eliminate overfitting.

However, overfitting and dataset biases do not necessarily undermine the capacity of ML systems to follow our rules and participate in our practices – even if we often cannot control and do not fully know what exactly such systems learn. For the same also often applies to humans. Wittgenstein presents a thought experiment that can be used to illustrate this. This thought experiment (PI: §185) revolves around a pupil who learns to continue the following numerical series: "2, 4, 6, 8, 10, etc.". The pupil does some exercises, none of them going beyond 1000. However, when he is asked to continue the series beyond 1000, he writes "1000, 1004, 1008, 1012". That is because, instead of "+2", the pupil learned "+2 if <1000, +4 if ≥1000". The pupil did not learn what the teacher intended him to learn. Yet, if he had not been asked to continue the series beyond 1000, the teacher would not have noticed. For, as far as tests up to 1000 are concerned, his answers were correct. One cannot look inside the heads of other people. Moreover, one has only been exposed to a limited number of rule applications performed by others.

Does that mean that other humans do not necessarily follow the same rules as oneself? This conclusion rests on a misguided conception of rule-following: it conceives of "rules" as internal representations (or interpretations) that result from example-based learning and treats them as more basic than practice. As noted, Wittgenstein rejects this take on rule-following and treats rules as irreducible parts of practices. Consequently, overfitting and dataset biases do not *per se* prevent ML-driven systems from following the same rules humans follow, unless they impede a system's capacity to participate in human practices (just as some internal

representations impede a person's participation in a shared practice, as in Wittgenstein's example of series learning).

Peter Hacker provides further objections to the idea that machines can follow rules; in his view, "it makes no sense to talk of a machine *following* a rule" (2019: 106). As noted, Hacker holds that machines that appear to be rule-followers merely behave in accordance with rules. To make that point, Hacker argues that we only ascribe the ability to follow rules and participate in human practices to beings that are, in the context of our rule-governed practices, capable of activities such as "justifying, noticing mistakes and correcting them by reference to the relevant rule, criticizing deviations from the rule, and, if called upon, explaining an action as being in accord with the rule or teaching others what counts as following the rule" (2019: 106f.).[75] To respond to this objection, one can point to the circumstance that even human rule-followers are not always able to do the things Hacker considers essential for the ability to follow rules: a language-user can, for example, follow tense-related linguistic rules without being able to justify or explain why she uses the past perfect rather than the past perfect progressive. Wittgenstein himself presents a similar example when he talks about "[c]alculating prodigies who get the right answer but cannot say how" (PI: §236). The following question "Are we to say that they do not calculate?" (PI: §236) seems to be a rhetorical question. Or imagine that old Beethoven was not only deaf, but also blind and mute – yet, still able to write notes (but *only* notes) and, thus, to compose music. He would not be able to teach, criticise others, or explain his compositions. Nevertheless, since he can still compose, we would not deny that he can follow musical rules. Thus, appeals to the activities listed by Hacker fail to discriminate humans and machines with respect to the ability to follow rules.

To deny machines the status of rule-followers and practice participants, Hacker further argues that a "computer does not understand the results it types out, does not know what the

[75] Shanker (2002: 1–11) makes a similar point, using arguments that resemble those presented by Hacker.

symbols it displays mean, for it neither knows nor understands anything" (Hacker 1997: 53). As shown, propositional knowledge or explicit understanding of how to correctly apply a rule is not necessary for the ability to follow rules. However, Hacker can reply that he is speaking about a more basic kind of understanding that does not need to be explicit or in propositional form. In fact, Wittgenstein criticises mentalistic understandings of "understanding" which reduce "understanding" to mental representations such as mental images (Russell 1995: 173), "ideas" (Locke 1975: 406), or "interpretations" (Ogden & Richards 1923: 209f.). Accordingly, Wittgenstein rightly rejects attempt to reduce "understanding" to mental representations or processes (PI: §§363, 501-06). Rather, he conceives of "understanding" as an ability. In his words: the meaning of words like "understands" and "knows" is "closely related to that of 'can', 'is able to'" (PI: §150).[76]

As far as language-games are concerned, the ability to understand a linguistic utterance "manifests in how the hearer reacts to the utterance" (Glock 1996: 373). More generally, understanding manifests in a being's capacity to meaningfully react to (meaningful) contributions to a given practice. Accordingly, "justifying", "noticing mistakes", and the other above-mentioned things Hacker deems essential for rule-followers and practice participants can be understood as behavioural criteria indicating a rule-follower's capacity to understand. If one conceives of "understanding" as the ability to react meaningfully within the context of a practice, "understanding" and "rule-following" cannot be separated from each other. Both capacities are mutually dependent: without understanding, one cannot participate in a practice and follow its rules; without the ability to participate in a practice and follow its rules, one cannot understand (meaningful) contributions to this practice.

Hacker argues that machines do not display understanding by pointing to very simplistic machines such as "[c]alculating devices" (2019: 106). However, even when it comes to much

[76] Dennett also challenges the "idea of comprehension or understanding as a separate, stand-alone, mental marvel" and suggests, not unlike Wittgenstein, that "comprehension is composed of competences" (2017: 94).

more capable machines of the future, Hacker would deny them the ability to understand and follow rules. For Hacker, machines in general, by virtue of being machines, are incapable of understanding and rule-following. In his view, it is conceptually impossible for machines (even for machines of the future) to display understanding, follow rules, and participate in human practices. This brings us to the next type of objection.

5.4 Form of Life Objection

Some Wittgensteinian sceptics of AI creativity might still object that it is conceptually impossible for machines to participate in human practices, and, thus, impossible for them to follow their rules. As we have seen above, practices are part of a larger context Wittgenstein refers to as "forms of life". So, the Wittgensteinian critics might contend that machines cannot participate in our practices and follow their rules since they do not share our human form of life. (In line with what has been said in the previous section, this also means that machines lack the capacity for understanding.)

To problematise this critique, it is, first, helpful to highlight the vast variety of intra-human forms of life, of cultural differences, within the history of humankind. Compare, for example, the cultural differences between New Yorkers on the one side and an indigenous tribe in the Amazon rainforest that has never been exposed to the "modern" world on the other. These two groups live in very different environments, have very different conventions, and, in part, also different desires. Unsurprisingly, both groups also have vastly different practices – and their abilities to participate in even the more mundane practices of the other group are very limited and, in many cases, absent. This raises the question of which feature of the human form of life all human rule-followers share, independent from the culture they belong to.

It might be tempting to suggest that, despite all these apparent differences, the inhabitants of New York and the Amazon River still share some basic desires, most notably, the desires to survive and to procreate. Machines, in contrast, seem to lack these desires.

However, there are asexual persons and people who commit suicide. So, not all humans share the desires for survival and procreation. Nonetheless, we would not say that people who do not have these desires are unable to participate in practices of the human form of life.

The Wittgensteinian gatekeepers might respond that New Yorkers, the members of the tribe in the Amazon, and people who lack the desires for survival and procreation still have in common that they are human beings (and as such capable of having "desires" or a "will" in the first place). So, they might contend that one needs to *biologically* be a human in order to be able to participate in human practices. After all, human practices are embedded in the *human* form of life. Yet, this overly anthropocentric view is also problematic. Consider, for example, the documented cases of "feral children" who are raised by wolves or primates. Because of significant biological or bodily differences, there is reason to assume that these children are not fully part of the wolf or primate form of life. Nonetheless, they can meaningfully participate in many practices of the wolf or primate form of life. The only way to deny this consists in resorting to essentialisms and dogmatisms which are irreconcilable with the late Wittgenstein's metaphilosophical views.[77] Obviously, bodily differences might hinder a "feral child" from participating in certain practices embedded in the wolf or primate form of life – but that does not mean that the child cannot participate in any of these practices. Likewise, intra-human differences (including disabilities) hinder some people from participating in practices that are "normal" to many other humans. Consequently, practice participation does not coincide conceptually with species boundaries. In principle, this is also true of human practices, even if, in practice, the fact that many human practices are based on or include human language renders it impossible for the non-human species we know of to participate in these practices.

[77] Resorting to metaphysical, rather than biological, conceptions of "the human" which emerge only in seventeenth-century Europe (Rees 2021) are no less essentialist and dogmatic – and by no means more compatible with Wittgenstein's metaphilosophy.

Wittgensteinian critics may react by shifting the emphasis from the *human* form of life to the human form of *life*. More concretely: in order to participate in human practices and, thus, to qualify as a rule-follower, one must be alive. (After all, despite the great plurality of intra-human forms of life, humans are united in the fact that they are alive.) This idea is also at the bottom of Hacker's critique: in his view, machines ultimately cannot do all the things he deems necessary for rule-following because they are not alive (1997: 55; 2019: 109ff.).

However, this line of argument is also problematic. First, there is no universally accepted philosophical or scientific definition of "life". Secondly, the view that "being alive" is necessary for rule-following rests on a biologistic bias. As mentioned above, in his critique, Hacker focuses on simplistic machines that perform "mechanistic" operations such as calculations. In other words, he keeps to a one-sided diet of examples when speaking of "machines". He avoids discussing more "capable" fictional machines we know from science fiction books and films such as C-3PO from the *Star Wars* franchise or HAL 9000 from *2001: A Space Odyssey*. In the *Star Wars* films, C-3PO can, *prima facie*, participate in human linguistic practices and, thus, follow the rules of language. We would also say that he can understand human language. In fact, his linguistic abilities are superhuman as he masters more than six million forms of communication (Grant & Reynolds 2016: 245). C-3PO is neither "alive", nor a biological being. Yet, given the linguistic performance displayed by C-3PO in the *Star Wars* films, it would be biologistic to deny him any competence to participate in human linguistic practices simply because he is an inanimate machine.

The Wittgensteinian critic might object that C-3PO is a *fictional* robot and that there are no machines that come even close to the linguistic abilities of an average human speaker, let alone C-3PO. That, however, does not mean that there will never be any machines with human-level or superhuman linguistic abilities in the future. In other words, such machines are not *conceptually* impossible. It may be possible that all future attempts to build such machines fail – but this cannot be shown on purely conceptual grounds. In contrast to Hacker's position, the

question about the future possibility of inanimate, non-biological machines with human-level linguistic competence is a meaningful question. It is a technical question, not a conceptually confused pseudo-question.

Hacker-inspired Wittgensteinians may appeal to the fact that C-3PO tends to behave like a living being (i.e., displays behavioural criteria associated with "being alive"). On this basis, they may propose that we only ascribe the ability to participate in human practices and follow their rules to beings that *behave like* a being that is alive. In this case, machines like C-3PO would be rule-followers. And one would not need to resort to the assumption that C-3PO and similar machines are actually alive – an assumption that would be confused: "Suppose we could construct a machine[…] with (nearly) all of the psychological attributes of persons: would [it] be alive? Probably not, given the nature of [its] hardware" (Luper 2019). By drawing on behavioural criteria, one can accept that C-3PO is a practice participant and rule-follower, without having to grant the same status to more simplistic machines. (Hacker, who uses the terms "life" and "animality" interchangeably, writes that "[w]hat [machines] lack is not computational power, but animality" (1997: 55; also see 2019: 111). Hacker is not just speaking of behavioural criteria linked to "animality" or "being alive", but of "animality" itself. So, he seems to assume that actual animality is necessary for exhibiting behavioural criteria related to "animality" or "being alive". However, this assumption is due to the one-sided diet of examples he keeps to when thinking about "machines".)

However, in some contexts, C-3PO does not behave as animals would behave. Unlike animals of his size, he shows, for example, no (criteria of) pain, e.g., when he loses a limb; he does not show any (criteria of) hunger or thirst; he is unaffected by heat and cold; he can survive and function normally in extraterrestrial environments. Thus, in many situations, C-3PO does not satisfy behavioural criteria associated with "being alive" or "animality". One might be tempted to respond that these situations are irrelevant. But this poses the question of how to

separate relevant from irrelevant situations. Declaring those situations relevant in which C-3PO's behaviour satisfies the relevant behavioural criteria begs the question.

Or think about HAL 9000 from *2001: A Space Odyssey* and imagine an earlier version of it, HAL 8000, if you will. HAL 8000 is like HAL 9000 as portrayed in Kubrick's adaptation, but has a mechanistic and monotonic, non-human-sounding voice and has not yet learnt to simulate human emotions. Given HAL 9000's disembodied form of existence, it is unlikely that many people would say that HAL 9000's behaviour displays "animality". It is even more unlikely that they would say that HAL 8000's behaviour satisfies the behavioural criteria associated with "animality". At the same time, it is unlikely that many people would say that HAL 8000 and HAL 9000 are incapable of understanding human linguistic utterances and communicating with humans (i.e., of participating in human language-games and following their rules). Thus, the attempt to restrict the ability to participate in human practices and follow their rules to machines that behave like animals or animate beings is, ultimately, unsuccessful.

Some Wittgensteinian gatekeepers may invoke Wittgenstein's famous statement that, "[i]f a lion could talk, we could not understand him" (PI: 225) that has already been referenced in a previous chapter. We cannot understand the lion because – even if he could talk in some lion language and, therefore, follow some rules – he is incapable of following the rules of *human* language-games. The Wittgensteinian gatekeepers might find it hard to accept that C-3PO and other machines can participate in human linguistic practices and follow their rules, while Wittgenstein denies lions the capacity to do so. But these doubts are due to biologistic biases. For there is a decisive difference between a lion and a speaking robot. Such a machine is built by humans for the purpose of communicating with humans (i.e., the purpose of participating in human language-games). If its linguistic capabilities are the result of some kind of machine learning, the data it has been trained on are texts which were written by humans. (OpenAI's GPT-3 which one of the most impressive language models today was trained on more than 45 terabytes of text.) In a sense, "technologies are always embedded in forms of life"

(Coeckelbergh and Funk 2018: 166). They are "woven into the texture of everyday existence" (Winner 1986: 12). The same applies to a speaking robot like C-3PO. He is, at least to some degree, embedded in, or woven into, the human form of life. Therefore, his utterances are – unlike the roaring of a lion – not merely noise to us. They are intelligible in the context of human language-games.

In a final effort to deny machines the capacity to participate in practices, some Wittgensteinian sceptics might (in a somewhat Heideggerian way) argue that what Wittgenstein calls "form of life" ought to be primarily understood as a shared context or horizon of relevance. As human practices are part of the human form of life, they are imbued with relevance. It follows that, in order to be able to participate in human practices, it is necessary to be sensitive to the relevance of practices. It is tempting to believe that this is impossible for machines. Yet, this view is confused. To avoid misunderstandings, it is helpful to distinguish between two kinds of "relevance", namely the relevance *of* a practice within a wider context and relevance *within* a practice. The question of whether something can participate in a practice concerns the latter kind of relevance, i.e., the question of whether something can give relevant responses or perform relevant actions *within* a certain practice. The ability to be sensitive to relevance *within* a practice does not necessarily presuppose an understanding or awareness of the relevance *of* a practice within the wider context of human life. Consider the following questions to see this: what is the relevance of chess within human life? What is the relevance of art? What is the relevance of philosophy? Many people who are capable of participating in these practices do not have answers to these questions. (Some might even deem these practices irrelevant in the grand scheme of things).

Back to relevance *within* a practice. It is tempting to assume that being sensitive to relevance *within* a practice requires one to be able to have some mental or conscious awareness of relevance. However, this view is problematic. We have seen that the ability to follow rules does not presuppose being conscious or mentally aware of them. Accordingly, participating in

a practice does not necessarily require consciousness or mental awareness of which actions are meaningful within a practice. By extension, due to the intertwinement of meaning and relevance, sensitivity to relevance *within* a practice does not require one to be conscious or mentally aware of relevance. Furthermore, we must not fall back on conceiving of practice-immanent relevance in terms of mental representations. If a mental representation fixes that a certain action counts as relevant within a practice, we can question the relevance of this mental representation. Introducing another mental representation pushes the question only one stage back which results in an infinite regress. Consequently, as practice-immanent relevance does not rely on mental representations, the ability to be sensitive to relevance *within* a practice does not require one to have mental representations. Instead, in line with the deflationary outlook of our approach, sensitivity to practice-immanent relevance manifests (like the ability to understand) in the ability to give relevant responses or perform relevant actions within a certain practice. It manifests in the ability to meaningfully participate in a practice. C-3PO, for example, is able to do so in the context of our linguistic practices. Therefore, he is sensitive to relevance within the practice of language, even if he does not have any conscious awareness or mental representations of it.

It becomes evident that appeals to form of life fail to clearly separate beings that are capable of rule-following from beings that are not. In order to avoid falling back on anti-Wittgensteinian essentialisms, biologistic prejudgments, or mentalistic preconceptions, we should not focus on form of life as the key concept to assess rule-following abilities, but on *practices* themselves.

5.5 A Practice-Based Account of Artificial Creativity

It has been shown that, not just when it comes to the metaphysical and epistemological questions surrounding rule-following that have been discussed in section 3.3, but also when it comes to the attributional question of who or what can follow rules, practices serve as the basic

level of inquiry. As we have seen, the attacks against computational creativity that have been scrutinised rely on inappropriately restrictive and inflated notions of practice participation and rule-following. In the face of their shortcomings, we reject such conceptions of practice participation and replace them with a deflationary one. According to a deflationary take on practice participation, beings qualify as practice participants if their contributions to a practice are meaningful in the context of this practice and its rules. Accordingly, if a human being utters meaningless nonsense or makes illegitimate moves in a game (for example, due to the effect of a certain drug), we would deny that this person is following the rules of language or the game in question. If it seems like this person in incapable of making meaningful utterances or correct moves, we would not say that this person participates in the respective practices. Meaningless contributions to a practice lack practice-immanent relevance. Analogously, meaningful contributions to a practice exhibit practice-immanent relevance.

If a contribution to a practice is meaningful, it has to manifest this "in actual cases" (PI: §201). Moreover, what represents a meaningful and relevant contribution to a practice needs to be assessed from the inside of the respective practice, rather than from the outside. In actual situations, competent practice participants can normally separate meaningful from meaningless contributions to the practice they participate in. They do not need to appeal to factors external to the practice but can normally do so by being immersed in the practice. Meaningful contributions to a practice normally resemble other contributions to that practice. In some situations (for example, when confronted with instances of novelty-oriented creativity), there might be disagreements as to whether something is a meaningful contribution to a certain practice or not. Yet, normally, competent practice participants have no problem separating meaningful contributions from "contributions" that lack practice-specific meaning and practice-immanent relevance.

Based on this deflationary take on practice participation and rule-following, one can say that, in order to participate in a certain human practice and to follow its rules, a machine's

contributions must make sense in the context of this practice and exhibit practice-immanent relevance. It is important to highlight the plural of "contribution-s". For a single isolated meaningful contribution to a practice may be due to a lucky coincidence. However, if a machine coincidentally produces one single output that meaningfully contributes to an established practice, while all its other outputs are totally unintelligible, it would not participate in that practice – accordingly, it would not follow its rules, at least, not the rules of that practice. Such a singular meaningful "contribution" may be in accordance with a rule, even though the being that produced it did not actually follow it. Thus, to qualify as a rule-follower and practice participant, a being has to be able to participate in the relevant practice and follow its rules in a *reliable* way – and not just once in a while. (This is in line with Wittgenstein's remark that rule-following requires regularity, even though the latter is not sufficient for the former.)

Furthermore, a rule-follower needs not only be able to participate in a practice in a somewhat reliable manner, but also in a somewhat *autonomous* way. Accordingly, you would not say that your computer contributes to a language-game when receiving an email from a colleague. That is because the email has not been autonomously generated by your computer. Therefore, it is the colleague that participates in the (digitally mediated) language-game, not your computer (which only serves as a tool or medium of communication). The ability to autonomously contribute to a practice requires some "agency". We have seen in section 2.2 that it is questionable if creativity necessarily requires agency. However, in order to determine if a being qualifies as "creative", without relying on questionable views, it is necessary to show that this being exhibits some kind of agency.

In addition, a genuine rule-follower needs to be able to participate in a practice in a flexible manner that is adaptive to practice-immanent relevance. Imagine, for example, a machine that can only give three answers (e.g., "yes", "no", and "maybe"). The range of possible responses such a machine can give is highly restricted (even if it is able to give these answers in a correct and reliable way and without the assistance of humans). In other words:

such a system is very inflexible in the responses it can give. Thus, its ability to participate in human language-games is extremely limited. Not all practices demand the same amount of adaptive flexibility to participate in them. In chess, for example, the scope of meaningful, rule-governed moves is more restricted than in our linguistic practices; accordingly, the demand for flexibility is less distinct in chess, but not absent. The challenges involved in building machines that are adaptive to practice-immanent relevance relate to the so-called "frame problem" or, since there are various different definitions of it, a set of problems discussed under this label, namely: how to build systems that can "think and act in ways that are adaptively sensitive to context-dependent relevance" (Wheeler 2008: 323), more specifically, how to build "a purely mechanistic system [that] might achieve appropriate, flexible and fluid action *within* a context" (Wheeler 2008: 340) which means – on our approach – within the context of a practice.

To sum up: genuine rule-followers and practice participants can meaningfully contribute to rule-governed human practices in a (i) reliable, (ii) autonomous, and (ii) adaptively flexible manner. This raises the question of whether machines qualify as rule-followers and practice participants on this basis. Yet, this is an ill-posed question. It is a misconception to assume that the ability to participate in human practices and to follow their rules is something machines, qua machines, possess or lack. As mentioned above, the ability to partake in a practice and to follow its rules has to be manifest in actual cases. Accordingly, whether *a* machine can participate in a certain practice and follow its rules, must be assessed on a case-by-case basis, for each system individually. It has to be assessed by looking at a machine's actual performance in the context of the relevant practice.

Above, we took it for granted that C-3PO can participate in human language-games. Is this assumption valid? Well, if you are sceptical, watch the *Star Wars* movies and observe how C-3PO linguistically interacts with human speakers – "don't think, but look!" (PI: §66). C-3PO reliably produces utterances that are meaningful in the context of human linguistic practices

and display a great amount of flexibility. So, C-3PO's linguistic performance satisfies the requirements for reliability and flexibility. But how about autonomy and agency?

As we have seen in section 2.2, there are attempts to reductively explain agency in terms of mental states and events such as an agent's intentions, desires, or beliefs. Since such mental states are usually viewed as having representational contents, this approach tries to explain agency in terms of mental representations (Schlosser 2019). However, it has been shown that one cannot explain rule-following in terms of mental representations. Thus, when thinking about agency – or, more specifically, the kind of agency that is required to follow rules –, we should abstain from explanations that rely on mental representations. For the same reason, we should refrain from embracing dispositionalist accounts of agency that strive to explain agency in terms of neurophysiological mechanisms or other sorts of dispositions.

As mentioned above, Wittgenstein emphasises that the ascription of allegedly inner phenomena requires criteria that are manifest in actual, context-specific behaviour (PI: §580). As far as the attribution of agency is concerned, Floridi and Sanders (2004) advance a similar idea which challenges mentalistic and neurocentric conceptions of agency and autonomy. They stress that agency is always ascribed to a being at a certain *level of abstraction*. The *level of abstraction* determines which information about a being is considered when making judgments about its agency. When attributing agency or autonomy to fellow human beings, we do not do so because we know their mental history or have neuroscientific evidence that the actions of humans are *not* completely predetermined by the laws of nature, the architecture of the human brain, the actions of other people, or environmental factors – since we, obviously, have no such evidence. Nonetheless, we commonly attribute agency and autonomy to our fellow human beings when interacting with them in ordinary situations. That is because we do so on a *level of abstraction* that takes information about their situated behaviour into account, rather than potential scientific insights or philosophical speculations about what ultimately causes human action. So, the question of whether agency and autonomy can be ascribed to machines ought to

be evaluated on the same *level of abstraction* on which we ordinarily ascribe agency and autonomy to other human beings. That means that machine agency or autonomy also has to be manifest in a system's performance.

We ascribe agency and autonomy, thus understood, to C-3PO. Sceptics may object that C-3PO has been built by engineers. However, C-3PO's engineers have no control over the exact linguistic outputs he will produce and cannot fully predict them. If the behaviour of a being is totally predictable (and if this circumstance is not due to an active decision on the part of this being to act in such a highly predicable way), it is normally not considered autonomous. Analogously, the less predictable the performance of a system, the more agency and autonomy we often ascribe to it, given that its performance is still meaningful and relevant. As far as current AI systems are concerned, programmers have (as noted above) often no idea which exact outputs or actions the system they built will produce or perform. Accordingly, they exhibit, at least, some amount of agency and autonomy as conceived above.

It has been shown that C-3PO can reliably, flexibly, and autonomously contribute to human linguistic practices. Gatekeepers might critically remark that C-3PO's ability to contribute to some language-games, for example, those revolving around human emotions, is somewhat constrained. However, no human is able to meaningfully contribute to *all* conceivable language-games. How many humans do you know who are able to competently contribute to language-games about quantum computing, medieval Japanese literature, ornithology, and numismatics? More generally, no human can participate in all existing human practices. It is, for example, questionable if there has ever been a human being capable of aptly contributing to the practices of astrophysics, opera singing, aerobatics, kung fu, contortion, and tightrope walking. It would be unfair to demand from a machine to be able to participate in all conceivable practices in a comprehensive way in order to accept that it can participate in some practices.

As C-3PO can reliably, flexibility, and autonomously contribute to human language-games in a meaningful manner, there is no reason to deny him (linguistic) creativity. Yet, C-3PO is the product of science fiction. So, how about cutting-edge AI systems today? In what sense can they be creative?

5.6 AI Creativity Today – Possibilities and Limitations

The chatbots and text-generators that exist today are less capable than C-3PO when it comes to participating in human linguistic practices. Current autoregressive language models such as OpenAI's GPT-3 and its successor ChatGPT may come closest to the linguistic performance of the robots we encounter in science fiction, but that does not mean that they are particularly close as they exhibit significant shortcomings: "GPT-3 makes silly errors that no human would ever make" (Rees 2022: 168).

This shows that that GPT-3's and ChatGPT's abilities to participate in human linguistic practices are more limited than that of C-3PO or human beings. Outside of natural language processing, many of today's AI systems exhibit similar limitations. But does that mean that these machines are incapable of participating in human practices and following their rules? This conclusion is too radical since, even among humans, the ability to participate in human practices comes in degrees. The average 40-year-old language user is, for example, normally more capable of participating in our linguistic practices than the average 4-year-old child. (And even among 40-year-olds, linguistic competence can vary significantly.) Does that mean that the average 4-year-old child is incapable of participating in human language-games? No, it only means that the former is more capable of doing so.

A toddler and beginner-level language learner usually make more mistakes than a competent native speaker. Furthermore, the latter is normally also more autonomous, i.e., less dependent on the assistance of more competent speakers, when participating in a language-game. In addition, the native speaker's utterances will be more flexible, varied, and attuned to

the specific context and situation. Thus, normally, the greater the ability to participate in human practices, the greater the relevant being's reliability, agency, and flexibility in meaningfully doing so.

When compared to older systems, GPT-3 exhibits a remarkable degree of reliability, autonomy, and flexibility in producing its linguistic outputs. However, many of its most impressive demos were cherry-picked from numerous trials (Lee & Qiufan 2021: 114). This shows that there are limitations as far its reliability is concerned. Moreover, as GPT-3 lacks the ability to perceive its environment, its capacity to participate in language-games that revolve around immediate sensory experience is also restricted. More generally, GPT-3 is a disembodied system. These factors limit its adaptive flexibility. Nevertheless, it would be unfair to deny systems like GPT-3 and its successor ChatGPT the ability to participate in human language-games altogether. Yet, it should have become clear that their capacity to do so is often limited. Similar things can be said about the majority of AI systems today when assessing their capacities to contribute to the practices their outputs relate to – they lag behind humans (as well as certain science fiction machines) in their abilities to do so. That does not necessarily mean that it is conceptually impossible for machines to reach human-level or even superhuman abilities to participate in human practices. It only means that this is – at the moment – practically or, if you will, technically impossible.

This raises the question of whether machines with limited abilities to participate in human practices can still be creative. To give a short answer: well, in a sense – but in a limited one. As highlighted, GPT-3, ChatGPT, and other existing language models can partake in human language-games only in a restricted way that exhibits limitations of reliability, agency, and adaptive flexibility when compared to most humans. However, the systems in question are normally still able to produce many meaningful outputs that are not predictable by their creators or interlocutors but attuned to the current language-game. Therefore, these systems exhibit *some* degree of reliability, agency, and adaptive flexibility in producing meaningful outputs. Hence,

when such a system produces creative outputs (e.g., a poem or joke), it would be unfair to deny it any creativity. At the same time, in the face of the above-mentioned limitations, one would go too far if one attributed human-level creative powers to them. As we will see shortly, similar things can be said about AI systems whose outputs relate to other practices. Another limitation of many existing AI systems is that they can perform only one task or a small number of tasks within a given practice. In other words, they can only follow a limited number of rules within a practice. It is often not possible for them to adapt to a new task or follow new rules within the same practice, let alone contribute to another practice.[78] And even if AI systems could accomplish this, it would often be difficult for them to identify which practice or which tasks or rules within a practice are relevant in a certain context or to achieve a certain goal.[79] Therefore, current AI systems need to be put in the right context by humans to perform their tasks.

For all these reasons, we would not regard most existing AI systems as full-blown participants in the practices their outputs or actions relate to. Therefore, we would not consider them full-blown creative agents. Nonetheless, some of their outputs or actions still strike us as creative within the narrow domain they contribute to. So, in a sense, such AI systems display "glimpses" of creativity. A being displays "glimpses" of creativity if its outputs or actions are "creative", although the being can only participate in the relevant practice these outputs or

[78] As far as neural networks are concerned, these limitations are, in part, also due to the so-called "problem of catastrophic forgetting" that was first described by McCloskey and Cohen (1989) and refers to the tendency of neural networks to forget all their old tasks once they learn a new one. Another problem in this context are the limited extrapolative abilities of neural networks and machine learning systems more generally. Mathematicians distinguish between "interpolation" and "extrapolation". The former stands for making predictions within the range of data it has been trained on. "Extrapolation", on the other hand, means making predictions that extend beyond the range of the training data. While machine learning systems excel at interpolation, their extrapolative abilities often fall short; in fact, it is questionable if it is possible to develop an extrapolative algorithm (Ye 2020). This limits the ability of a machine learning system to perform tasks that go beyond the narrow scope of what it has been trained to do, both within the same practice of the task it masters and in the context of other practices.

[79] The concerns that are raised in this paragraph relate to the frame problem, more specifically, to two aspects of the frame problem. In allusion to a distinction introduced by Wheeler (2008: 340), one can speak of an "intra-practice frame problem" and an "inter-practice frame problem". The former centres on how to achieve flexibly adaptive action within a practice; the latter revolves around how to identify the relevant practice and switch between practices if necessary.

actions are part of in a limited manner, and so does not qualify as a full-blown creative practice participant.

In *some* respects, the ways in which many AI systems, i.e., our digital brainchildren, can participate in established practices resembles the ways in which (human) children can do so. If the latter are capable of participating in practices, then often also only in a way that exhibits limited reliability, autonomy, and flexibility in bringing about meaningful contributions. Moreover, like AIs, children often need to be put into the right context and require (human) assistance in order to participate in given practices and deliver creative contributions. Even if this matter cannot be adequately addressed here, I think the notion of "glimpses" of creativity is, in many cases at least, also useful when thinking about the attribution of "creativity" to children.

We have so far focused on language models. Let us turn to an AI-driven system that attracted a significant amount of attention, both in technical and artistic circles: this system was developed by a group of researchers at the University of Tübingen (Gatys et al. 2015) and is, through the power of deep learning, capable of imitating virtually any painterly style that has been developed within the practice of painting, including the distinct styles of Turner, Van Gogh, Munch, Kandinsky, to name but a few. However, even if the system can somewhat reliably, autonomously, and flexibly mimic existing artistic styles, its ability to participate in the human practices of painting and art more generally is very limited. Nonetheless, it displays glimpses of value-oriented creativity. Similar things can be said about deep learning models such as OpenAI's DALL-E and its successors (which are based on GPT-3). These models can generate new images (that are not contained in the dataset they have been trained on) upon receiving so-called "text prompts", i.e., natural-language descriptions of the desired content of the image.

The same assessment seems appropriate with regards to a model like DeepBach (Hadjeres et al. 2017) which can (somewhat reliably, autonomously, and flexibly) generate

chorales in the style of Johann Sebastian Bach. In a Turing-Test-like experiment, volunteers had to separate DeepBach-generated pieces from chorales composed by Bach himself. Astonishingly, in 50% of the cases, DeepBach's outputs were taken to be actual Bach compositions. The composition students and professional musicians that were part of the study were only marginally better as they considered DeepBach's creations actual Bach chorales in 45% of the cases. This is an astounding result, especially if one considers that "[c]horales are pretty unforgiving" so that it "takes just one bum note to out one as an imposter" (du Sautoy 2019: 197). However, despite DeepBach's impressive ability to imitate the style of Bach's chorales in a relatively reliable, autonomous, and flexible way, its capacity to contribute to the practice of music beyond that is extremely limited. So, it too only provides glimpses of value-oriented creativity.

Another interesting example of "artistic" AIs, so to speak, are so-called Creative Adversarial Networks or CANs (Elgammal et al. 2017: 4). CANs rest on a ML framework called Generative Adversarial Networks or GANs which consist of two neural networks, a "generator" and a "discriminator". CANs also feature a generator and a discriminator. The generator of a CAN, which has no access to any artistic data (i.e., digitised paintings), generates images starting from a random input. The CAN discriminator, in contrast, has access to a huge set of labelled artistic data, with the labels specifying the style of the featured artworks. For any image the generator generates, it receives two signals from the discriminator: the first signal indicates whether the discriminator considers the presented image "art" or not; technically put, it signals the discriminator's assessment of whether the image is coming from the same distribution as the data it has been trained on. The second signal indicates how easily the discriminator can classify the image into the artistic styles it has learnt. In contrast to the generator of normal GANs (that tries to convince the discriminator that the images it generates are from the same distribution as the data the latter has been trained on), the generator of a CAN tries to generate images the discriminator classifies as "art" but, at the same time, fails to clearly

classify into existing styles. Less technically speaking: CANs try to "increase the stylistic ambiguity and deviations from style norms, while at the same time, avoiding moving too far away from what is accepted as art" (Elgammal et al. 2017: 4f.).

In a Turing-Test-inspired experiment, a pool of non-expert subjects was shown CAN-generated images alongside paintings that were exhibited at Art Basel 2016, without knowing which of the images were created by a computer. Surprisingly, the Art Basel works were, on average, more often considered computer-generated than the CAN outputs (Elgammal et al. 2017: 15f.). In fact, more than half of the CAN images were judged to be human works of art. In another experiment, a pool of art history students was shown pairs of images: one of the images was generated by a CAN, the other one by a variant of CAN (sc-CAN) that does not try to maximise stylistic ambiguity so that it merely mimics the styles of the artworks that are part of the data set the discriminator has access to. CAN images were considered more aesthetically pleasing by 60% of the subjects and more novel by roughly the same proportion than the sc-CAN images. For the reasons that have been invoked with regards to the above-mentioned models, the participatory capacities of CANs are limited in the context of human artistic practices. So, they also offer us only glimpses of creativity. However, the results of the above-mentioned experiments can be interpreted in a way that suggests that CANs provide glimpses of *sweet-spot,* rather than value-oriented, creativity – to be more precise, glimpses of non-cutting-edge sweet-spot creativity. For, due to the architecture of CANs, the novelty their images exhibit is somewhat confined, as the model "tries to generate art that is novel, but not too novel" (Elgammal et al. 2017: 4).

The question that looms over this discussion is whether there are any existing AI systems that are capable of more than "glimpses" of creativity. Are there any AIs that can fairly comprehensively participate in an established human practice and produce creative contributions to this practice? AIs have always excelled in producing contributions to relatively well-structured practices with comparatively strict or "algorithmic" rules. Computers can, for

example, easily participate in the practice of calculating – even with higher reliability and at a faster pace than humans. This is underscored by the Church-Turing-Thesis according to which every effectively calculable function is Turing-machine-calculable. In other words, all algorithms can be represented in binary terms so that they can be computed by a Turing machine. Consequently, a Turing machine can compute everything a human "computer" can compute by following the fixed rules of an algorithm. If one accepts the Church-Turing-Thesis (as most computer scientist and mathematicians do), computers can follow *all* the rules of calculation we humans can follow and, thus, fully contribute to the practice of calculating. In this sense, computers are "rule-following beasts" (Hofstadter 1980: 26).

However, there is little space for creativity in calculating. In the wider practice of mathematics beyond calculations there is plenty of space for creativity, but the way in which current AI systems can contribute to higher mathematics is still limited.[80] However, there are other practices AIs can participate in in a somewhat comprehensive way, namely, games. Deep Blue and AlphaGo can, for example, follow the rules of chess and the ancient Chinese game of Go in a comprehensive way and, thus, fully participate in these practices. Since the rules of chess are comparatively strict and "algorithmic", brute computational force and logic were enough to make Deep Blue participate in the practice of chess and beat Kasparov, the best human practitioner of chess at the time. That, however, is no reason to deny Deep Blue the ability to follow the rules of chess: again, it is not internal processes, but actual performance, that determines if a being can participate in a practice and follow its rules. It may also be tempting to deny AlphaGo the ability to participate in the practice of Go because it requires the help of a human to move the stones on the board. However, this would be like denying Stephen Hawking the ability to participate in human linguistic practices because he relied on speech-generating technologies to do so. Andy Okun, president of the American Go Association – and,

[80] Some limitations are discussed by du Sautoy (2019: 233-53).

thus, a very competent participant in the practice of Go –, describes the situation by comparing the practice of Go with a conversation: "The time when humans can have meaningful conversations with an AI has always seemed far off and the stuff of science fiction. But for Go players that day is here" (quoted from Miller 2019: 52). Similarly, Kasparov's initial reaction to Deep Blue was that he accused IBM of cheating: he could only make sense of Deep Blue's performance by assuming human intervention behind it (Miller 2019: 46). In line with that, while we probably would not say that DeepBach "composes" or CANs "create" art (without using inverted commas), we – including competent participants in the practice of Go – tend to be less hesitant to say that AlphaGo plays Go and Deep Blue plays chess.

Some gatekeepers may deny chess computers and AlphaGo the capacity to participate in the practices of chess and Go by arguing that these machines do not show any excitement, frustration, or surprise. Moreover, they may point to the circumstance that they cannot decide *not* to play chess or Go. This objection, however, rests on an overly inflated conception of what it takes to participate in a practice. The emotions a practice participant displays while participating in a practice can be quite arbitrary as far as many practices are concerned. Imagine a Zen master who does not display any emotions (and may not even feel any emotions) while playing chess. Imagine further that the Zen master plays against a person who is, due to a spell or a strong psychological disposition, forced to play chess, whenever he gets the opportunity to do so, and is dispositionally incapable of stopping a chess match. (In *Schachnovelle*, Stefan Zweig arguably describes a similar pathology, an extreme "compulsion to play [chess], a mania, a frenzy" (2006: 63); with this pathology, this "chess sickness", it "would be better not to go near a chessboard" (2006: 73) as one would not be able to stop playing by one's own efforts.) Would we say that the Zen master and the "chess sick" opponent participate in the practice of chess? Well, yes, if the moves they make are correct, manifest an understanding of the game, and exhibit practice-immanent relevance, i.e., if the moves they make are meaningful.

The victory of DeepMind's AlphaGo over 18-time world champion Lee Sedol, one of the greatest (human) Go players of all time, in a five-game match in 2016 was undoubtably one of the most shocking moments in AI history so far. Since there are significantly more possible board configurations in Go than atoms in the universe, computing power and logic are not enough to make an AI system successfully participate in the ancient game of Go. Therefore, the developers of AlphaGo combined GOFAI methods with reinforcement learning, i.e., a technique that makes neural networks learn by rewarding certain behaviours. AlphaGo could not only play Go in a reliable, autonomous, and adaptively flexible way – to the great surprise of many experts, it won four out of the five matches against Lee Sedol. In doing so, it not only displayed sweet-spot, but even *cutting-edge* creativity. Especially the 37th move of the second game attracted a lot of attention in the Go world. "Move 37", as it is commonly referred to, was so uncommon and counterintuitive that commentators thought that AlphaGo had made a grave mistake which revealed severe limitations of the system. However, about 50 moves later, AlphaGo emerged victorious – and it was move 37 that had sealed the victory. Over the course of these 50 moves, move 37 changed from being considered an error or, at most, an instance of novelty-oriented creativity to being appreciated as an instance of cutting-edge creativity, as "one of the most creative [moves] in Go history" (Tegmark 2018: 88). As Go champion Ke Jie puts it: "Humanity has played Go for thousands of years, and yet, as AI has shown us, we have not yet even scratched the surface" (quoted from du Sautoy 2019: 39).

AlphaGo's style and strategies (like the one underlying move 37) have eagerly been adopted and imitated by human players. So, AlphaGo has changed how humans play Go today. Consequently, DeepMind's masterpiece displayed *game-changing* creativity – and a radical kind of game-changing creativity on top as it has transformed the practice of Go in a fundamental way. Similar things can be said about IBM's Deep Blue and other chess computers. Kasparov describes the situation as follows: due to the fact that virtually all professional chess

players today train with chess computers, their style tends to "reflect much silicon influence" (Kasparov & Greengard 2018: 50).

As AlphaGo is capable of participating in the practice of Go in a reliable, autonomous, and flexibly adaptive manner, there is a stronger tendency to ascribe "creativity" not just to its moves, i.e., its outputs and actions, but to the system itself. In other words: as AlphaGo is a competent participant in the practice of Go, we more readily call *it* "creative" than systems whose ability to participate in their respective practices is more limited. Accordingly, after the match against AlphaGo, Lee Sedol stated: "Surely AlphaGo is creative" (quoted from Leach 2022: 2). And the Go world agreed, including previous sceptics and gatekeepers.

So far, the number of practices AI can meaningfully participate in in a somewhat comprehensive way is still very limited. These practices tend to be clearly structured by somewhat strict rules. Moreover, they feature a clear measure of success (such as "winning" in the case of chess and Go) which allows for a precise formulation of the reward function. When it comes to practices that lack these features, it is more difficult to reach human-level participatory abilities. In the context of these more "AI-resistant" practices, AIs are still more like tools rather than full-blown practice participants. We have mainly focused on artistic practices here, but the same also applies to scientific practices. Consider, for example, DeepMind's AlphaFold which is widely considered AI's greatest success story in the context of science. AlphaFold is a deep-learning-based system that is capable of predicting 3D models of protein structures with high accuracy – a process that, if done experimentally, is highly expensive and lengthy. Even if AlphaFold is much faster, less expensive, and more scalable than more traditional methods, it is, at the end of the day, a tool (for protein structure prediction) rather than a full-blown participant in the practice of science.

However, over the last decades, AIs have become increasingly capable of making increasingly meaningful contributions to an increasing number of practices in an increasingly reliable, autonomous, and flexibly adaptive way. The examples that have been discussed

confirm this. Accordingly, resistance to viewing AIs as creative agents becomes more and more fragile in more and more practices. The circumstance that AI-generated pictures win art prizes (Roose 2022) and are offered by leading auction houses (Cohn 2018) – as well as the highly emotional reactions such controversial decisions trigger – illustrate this vividly. Of course, many people disagree with these decisions and reject that idea that current AI systems can be creative. However, the circumstance that there is disagreement about this matter in the first place would have been unthinkable decades ago. This demonstrates that AI systems have become more creative over the years – even if most of them are currently not able to display more than glimpses of creativity.

The above-mentioned AI systems displayed (glimpses of) value-oriented and game-playing as well as (glimpses of) sweet-spot and game-changing creativity. Moreover, the CAN-generated images the subjects participating in the experiment were divided about whether they represent "art" or not can be regarded glimpses of novelty-oriented creativity. But what about game-inventing creativity? This diachronic kind of creativity is more difficult (if not impossible) to achieve for current AI systems. Inventing a new practice often requires that one asks the right questions: how can be overcome this problem? How can this need be satisfied? What if this or that was the case? Yet, today's AIs fails to ask questions, let alone the right ones. In the words of Kasparov: "Computers are excellent tools for producing answers, but they don't know how to ask questions, at least not in the sense humans do […]. But to know which questions are the right questions, you have to know what is important, what *matters*" (Kasparov & Greengard 2018: 70, 72). As the extent to which current AI systems can be immersed into the human form of life which gives relevance to human practices is limited, these machines are not in the position to ask the right questions and to know what matters (to humans). Thus, accomplishing game-inventing creativity (or glimpses thereof) represents a major obstacle for AIs today (even if it is not entirely impossible that this obstacle will be overcome in the future).

Finally, one might wonder how to make AIs more creative (in line with the Paradigm 2 notion of "creativity"). To do so, we need to increase the creative agency of AI systems. This, in turn, requires us to turn AIs into more competent practice participants and rule-followers. In other words, we have to increase their reliability, autonomy, and adaptive flexibility in producing meaningful contributions to human practices. Maybe some of the features AI researchers deem necessary for engineering human-level AI – such as common sense (Shanahan 2015: 6; Choi 2022)[81], causal understanding (Pearl & Mackenzie 2018), and abductive reasoning skills, i.e., the ability to make inferences to the best explanation (Larson 2021), or situated embodiment (Brooks 2002; Pfeifer & Bongard 2007) – will help us to turn AIs into more competent practice participants and rule-followers in a wide variety of practices. Moreover, features like phenomenal consciousness or imagination (mentalistically construed as the capacity to produce inner representations that go beyond past and current perceptions) may help to increase the creativity of the contributions such practice-participating AIs can make to practices. (This does not imply that these mental phenomena are *conceptually* necessary for "creativity". A good ear, for example, is helpful, but not conceptually necessary, for becoming a world-class piano player.)

However, what it takes to meaningfully participate in a practice in a reliable, autonomous, and adaptively flexible manner ultimately depends on the individual practice and its rules: consider, for example, what it takes to successfully participate in chess as compared to the practices of opera singing or tightrope walking. It remains to be seen if AIs will succeed in developing the ability to participate in virtually all human practices, let alone in creatively doing so. Hans Moravec metaphorically describes "a 'landscape of human competence,' having lowlands with labels like 'arithmetic'" and "foothills like 'theorem proving" and 'chess

[81] Common sense is often construed in mentalistic terms. However, it is not necessary to ground common sense in mental representations. For one can conceive of common sense as something that "will be manifest in behavior. Or more likely, the lack of some aspect of common sense will be manifest in behavior" (Shanahan 2015: 6).

playing,'" (1999: 70) and Go (as we may add). Moravec compares advances in AI development to "water slowly flooding the landscape. A half century ago it began to drown the lowlands, driving out human calculators and record clerks, but leaving most of us dry. Now the flood has reached the foothills" (1999: 70). There are still mountain peaks that have not (yet) been flooded, including practices like "art" and "science" which are widely considered the highest mountain tops (Tegmark 2018: 80). The future will show if these mountain peaks will be fully submerged or defy the flood. If it turns out to be the former scenario, the next Picasso and the next Marie Curie may be AIs.

6. Coda: A Posthuman Paradigm of Creativity?

In the previous chapter, it has been shown that it is conceptually possible for machines to be creative, even if most of today's AI systems are not capable of displaying more than mere glimpses of creativity. However, we were tackling this question against the backdrop of Paradigm 2, that is, with respect to creativity as paradigmatically exhibited by humans. In other words: we conceived of creativity in relation to the human form of life or (less anthropocentrically put) the human form of being. Yet, as the transition from Paradigm 1 to 2 demonstrates, the reign of the human-centred Paradigm 2 is not set in stone. It can be superseded by another paradigm. Indeed, already in the nineteenth century, one can spot invocations of a new paradigm, of Paradigm 3, if you will. Paradigm 3 thinking abandons the idea that humanity represents the peak of creativity. Instead, it envisions *posthuman* creativity as the highest form of creativity.

In the nineteenth century, visions of such beings were chiefly inspired by Charles Darwin's *The Origin of Species*. Francis Galton, Charles Darwin's half-cousin, for example, drew on his half-cousin's theory of evolution to explore the future of genius. In his 1869 book *Hereditary Genius*, Galton writes that, "as it is easy[…] to obtain by careful selection a permanent breed of dogs or horses gifted with peculiar powers […], so it would be quite practicable to produce a highly-gifted race of men by judicious marriages during several consecutive generations" (1892: 1). As Galton deemed this not only "practicable" but also desirable (1892: 332), he was one of the first advocates of eugenics; in fact, the term "eugenics" is his coinage (1883). Galton himself mainly speaks about producing a population of "men of a high type" (1883: 44), rather than *post*humans. Other eugenicists like Julian Huxley, however, conjure "a new kind of existence, as different from ours as ours is from that of Pekin man" (1957: 17). So, the eugenic movement (explicitly or implicitly) points to a future with posthuman beings that possess superhuman creative abilities.

One of the earliest philosophical invocations of a posthuman paradigm can be found in the later writings of Friedrich Nietzsche. In his first book, *The Birth of Tragedy*, Nietzsche had argued that great art results from a fusion of two complementary metaphysical forces which he dubs the *Apollonian* and the *Dionysian*: "the supreme goal of […] all art" can only be attained through "a bond of brotherhood between the two deities" (1999: 104). In doing so, the early Nietzsche paints a picture of "suprahuman genius" (Andreas-Salomé 2001: 91). By contrast, in the positivistic middle phase of his intellectual development, Nietzsche sets out to demystify supernatural takes on artistic creation. Accordingly, he reinterprets art as a product of the "efficient workman" (Nietzsche 1996: 86) and reconceives genius as something human, all too human. So, while Nietzsche's early vision of artistic creation is indebted to Paradigm 1, his second account stands in the tradition of Paradigm 2 thinking.

In the final stage of Nietzsche's intellectual career, however, his Paradigm-2-inspired perspective gives way to another vision that culminates in one of his most controversial ideas, namely the "Übermensch", the "overhuman". Clearly influenced by Darwin, Nietzsche writes in *Thus Spoke Zarathustra* that "[m]ankind is a rope fastened between animal and over[hu]man" (2006: 7). Even if there are exegetical disagreements about how he envisions overhumans (Schrift 2001: 58), it is uncontroversial that they possess supreme creative powers. Accordingly, they are able to "create new values" (2006: 17) that overcome Christian values as well as the looming threat of rampant nihilism that results from Nietzsche's oft-quoted diagnosis of the death of God: "Dead are all gods: now we want the over[hu]man to live" (2006: 59). Despite Darwin's apparent influence, Nietzsche (2005: 101) rejects Darwinian interpretations of his concept of the *Übermensch* which conceive of the overhuman as a new biological species. It is still humans that turn into overhumans: humans need to "create over and beyond [themselves]" (Nietzsche 2006: 48) to become overhuman and reach unknown creative powers. Yet, instead of eugenic methods, Nietzsche seems to regard relentless self-realisation, "free[ing oneself]

from the morality of custom", and an "independent, enduring will" (Nietzsche 2007: 37) as pathways to achieve this.

Nowadays, (traditionally) eugenic ideas and Nietzsche's vision of the *Übermensch* are, for good reason, out of fashion. Today, we know that Galton's views on the inheritance of genius rely on an incorrect, pre-Mendelian account of inheritance. Moreover, eugenics usually went hand in hand with racism. Galton, for example, believed that Europeans possess superior mental abilities compared to other races (1892: x). In the 1930s and 1940s, eugenic ideas were adopted by the Nazis who established eugenicist and racial hygiene programmes millions of innocent people fell victim to. Nietzsche's notion of the *Übermensch* was instrumentalised by the Nazis for the same purposes.

However, there is an alternative vision of posthuman creativity. Again, Darwin serves as a key source of inspiration – for this third vision arguably makes its first appearance in the 1863 essay "Darwin Among the Machines" by Samuel Butler. The essay raises the question of "[w]hat sort of creature man's next successor in the supremacy of the earth is likely to be" (1923: 210). Not unlike the eugenicists and Nietzsche, Butler is convinced that "we are ourselves creating our own successors" (1923: 210). Yet, unlike the eugenicists and Nietzsche, he does not think that we will do so through selective breeding or radical self-realisation, but through "the vast development of the mechanical world" (Butler 1923: 208). Butler regards machines as the most likely contender for humanity's successor as the supreme being on earth. Thus, he paints a picture of the future in which "man will have become to the machine what the horse and the dog are to man" (Butler 1923: 211). More dramatically put: "In the course of ages we shall find ourselves the inferior race" (Butler 1923: 210). This implies that, in the future Butler envisions, machines will replace humans as the pinnacle of creativity.

Even though Butler's essay is largely neglected, it is arguably the first expression of what represents the most-talked-about vision of posthuman creativity today. As to the question of how to imagine the technologically created successor to *Homo sapiens*, there are currently

two main narratives. The protagonist of the first narrative is the "cyborg", i.e., a technologically enhanced human.[82] Due to the cyborg's allegedly superior powers, futurists and techno-utopians are convinced that its emergence "open[s] up new horizons of scarcely imaginable creativity" (Selinger 2009: 155). The idea of the cyborg is at the heart of a movement which attracts more and more supporters today, namely *Transhumanism*. Its main goal consists in "the continuation and acceleration of the evolution of intelligent life beyond its currently human form and human limitations" (More 2013a: 3). Transhumanists strive to achieve this with the help of present and potential future technologies, including classical cyborgian (mechatronics-based) technologies such as implants and brain-computer interfaces as well as biotechnologies such as nootropic drugs, bioartificial organ manufacturing, and hormonal and cellular manipulation. Transhumanism has been associated with eugenics (Rutherford 2022) and the Nietzschean *Übermensch* (Sorgner 2009, More 2010; More 2013a: 10); yet, what sets Transhumanism apart is that it tries to bring about posthuman beings through "technological transformation" (Bostrom 2005: 4). Since such posthumans possess radically enhanced capacities, it is assumed that they are capable of superhuman creativity: "These superhumans will enjoy unheard-of abilities and unprecedented creativity" (Harari 2015: 346). Once humanity has been "updated" to posthumanity, we are told, "entirely new possibilities of […] creative endeavor come into view" (Bostrom 2008: 132).

Most Transhumanists believe that our posthuman successors will evolve from us, from humans: "*we* can become posthuman" (More 2013a: 4; emphasis added). In other words, posthumanity will arise by "increasingly integrat[ing] our advancing technologies into our selves" (More 2013b: 450). However, there is a second narrative about how futurists conceive of the technologically created "successor" to humankind. The protagonist of this alternative narrative (which seems to be more closely aligned with the one originally told by Butler) is

[82] The term was first used in the context of the space travel debates of the 1960s with respect to the question of how astronauts can adapt to the hostile conditions of extraterrestrial environments (Clynes & Nathan 1960: 26).

created by, but does not biologically evolve from, humans: it is the machine itself – or, more precisely, a specific kind of machine – that will succeed and surpass humanity, namely a future form of artificial intelligence.

Many futurists believe that such an AI with superhuman abilities will be the result of an "intelligence explosion". This idea was introduced by I. J. Good who speculated that, once there is "an ultraintelligent machine […] that can far surpass all the intellectual activities of any man however clever", such a machine "could design even better machines" (1965: 33). In other words, a process of recursive self-improvement will make technology ever more powerful so that, eventually, "the intelligence of man would be left far behind" (Good 1965: 33). The anticipated point in this futuristic vision at which AI becomes so powerful that it significantly surpasses humankind is often called the "singularity".[83] In a word: "superhumanity is the essence of the Singularity" (Vinge 1993: 13). To support the idea of a looming intelligence explosion, the prophets of the singularity often invoke what is commonly called "Moore's Law". Following Moore (1975), this so-called law (which proved accurate for several decades) predicts that computing power doubles circa every two years, i.e., it increases exponentially. For Moravec (1988; 1999), Kurzweil (2004), and other futurologists, exponential growth is not restricted to computing power but applies to technological evolution more generally. Kurzweil even posits a "law of accelerating returns" according to which the returns of exponential technological advancement also improve exponentially so that there is "*exponential growth in the rate of exponential growth*" (2004: 381).

Many heralds of the singularity hold that AI will ultimately become capable of superhuman creativity: "AI will be more creative than humans" (Goertzel 2019). What is more, for many summoners of the singularity, an intelligence explosion seems to go hand in hand with

[83] The term "singularity" originally stems from general relativity theory, denoting the condition at the centre of a black hole where spacetime begins to break down. John von Neumann is usually credited for relating this notion to technological advancement (Ulam 1958: 5). For a philosophical take on the singularity, see Chalmers (2016).

a creativity explosion: even though technology might not reach "infinite creativity, […] it certainly moves rapidly in that direction" (Kurzweil 2010: 685). Therefore, this vision of the future entails "a feast of creativity resulting from human intelligence expanded manyfold" (Kurzweil 2010: 702).

While "the existence of the Creator depriv[ed] human beings of their creativity" (Mason 1988: 709) during the height of Paradigm 1 in ancient and medieval times, Paradigm 3 thinking does not deny that humans are capable of creativity. For it is human creativity that brings about posthuman beings with superhuman powers. In other words, such beings are humanity's invention or, as some argue, "the last invention that man need ever make" (Good 1965: 33). So, while humans are the creations of the supremely creative instance under Paradigm 1, they are its creators under Paradigm 3. Moreover, while Paradigm 1 thinking relied on a metaphysical worldview and a belief in the divine, Paradigm 3 thinking rests on a secular worldview and (in its techno-futurist form) an unshakeable belief in technological progress. Nevertheless, many prophets of Transhumanism and the singularity use pseudo-religious language when describing the technologically enhanced or produced superior posthuman beings at the centre of their vision of the future. Harari, for example, speaks of our "godlike descendants" that possess "divine powers of creation" (2015: 45f.).

If the techno-futurist picture turns out to be accurate and superhuman AIs or cyborgs will embody the new paradigm of creativity, it is likely that a conceptual shift in how we conceive of "creativity" will take place. This raises the question of how to conceive of "creativity" in times of a potential techno-futurist Paradigm 3. Just as Paradigm-2-based "creativity" is connected to the human form of being, "creativity" during Paradigm 3 will most likely be connected to the form of being of the creature that represents the peak of creativity. In the cyborgian version of the story, this will be the cyborg; in the singularity version, it will be an AI. To tackle the question of what "creativity" might look like in times of a posthuman Paradigm 3, without falling back on anthropocentric preconceptions, we would have to explore

what "creativity" that is true to the form of being of a cyborg or future AI would look like. As briefly mentioned above, some philosophers contend that we already are (Clark 2003) or have always been cyborgs (Sorgner 2021). Therefore, the cyborgian form of being seems to be not too radically different from the human one – at least, in comparison to the form of being of a future AI on which we will focus here. So, again: how to conceive of creativity in AI terms – in a way that is true to the form of being of such an AI?[84]

This is, of course, a highly speculative question, given that we have no idea what a future AI will look like. There is a wide space of possible AI forms of being. In fact, from our current position, this space is unfathomable, especially if future AI development follows Kurzweil's law of accelerating returns. This makes it impossible for us to make accurate predictions about the conceptual future of "creativity" during a potential AI-centred paradigm. Nevertheless, it is possible to highlight certain features that an AI with superhuman powers is, from the vantage point of today, likely to exhibit.

First, electronic processing is much faster than the processing involved in neural activities: "The orchestrated motion of electrons, which is at the heart of modern artificial information processing, can be much faster than the processes of diffusion and chemical change by which brains operate" (Wilczek 2020: 71). There are estimations that electronic processing is almost a billion times faster (Wilczek 2020: 71). However, there are physical limits on how fast and powerful a computer, and thus a computation-based AI, can be. The upper limit of a laptop-sized computer has been estimated to be 1051 operations per second (Lloyd 2000) which is around a billion trillion trillion times faster than the most powerful computer as of early 2019, which itself already marginally exceeded the raw capacity of the human brain (Russell 2019: 37). So, physics allows for machines billions of times faster than the human brain.

[84] For a similar suggestion in a different context, see Erden (2010: 361).

Moreover, in contrast to humans, AIs are not restricted to a single body: a "human has direct control over only one body, while a machine can control thousands or millions" (Russell 2019: 95). So, the pervasive tendency to imagine future AIs as single robots seems questionable. Instead, such AIs will most probably exhibit heavily distributed forms of embodiment. In that case, such an AI will be able to control a multitude of robots. Furthermore, it is likely that such a future AI will be powered by a network of distributed data centres with enormous storage and a colossal number of computing servers and processors. The massively distributed way in which the AI's memories and software are stored over multiple locations results in a great amount of robustness: if a single robot that is controlled by the AI or maybe even a whole data centre gets destroyed, the AI might still be fully operational. What is more, while humans need to rest, a future AI may be free from such demands and, thus, has higher duty cycle (Wilczek 2020: 71).

Due to these factors, AIs can be bolder and more daring than humans when pursuing creative efforts. Because of the availability of various (robotic) bodies and the possibility of backing themselves up, they do not need to care as much about physical integrity as humans do. Moreover, they can work at much higher speed and have a higher duty cycle than humans. Through the internet (or one of its successors), they will also have access to almost the entirety of documented human knowledge, including the majority of books ever written and all scientific discoveries made by humans. Therefore, they can draw on a tremendously wider knowledge basis than individual humans, encompassing oceans of knowledge produced by humans as well as other AIs. In addition, *prima facie*, the communication between different AIs seems likely to be more frictionless than between humans as the former can communicate over the internet and can read each other's code. In light of all these factors, it seems like the sheer size and scale of Paradigm 3 creativity would vastly exceed Paradigm 2 standards. Lastly, the space within which future AIs operate is not restricted to humanity's "terrestrial comfort zone" (Wilczek 2020: 75). For AIs "can survive under conditions that biological life struggles with, including

interstellar travel" (Schneider 2019: 11). Thus, the creativity of future AIs has the potential to be more spatially expansive than human creativity.

So far, we have mainly concentrated on probable bodily features of future AIs and, to a lesser degree, the environment they are likely to inhabit. But how about their desires or, if this strikes a too anthropocentric tone, their goals? Bostrom argues that future superintelligent AI could, in principle, feature "more or less any final goal" (2014: 130). The final goals of such an AI are, of course, hard to predict. However, Bostrom contends that even "without detailed knowledge of an agent's final goals, we may be able to infer something about its more immediate objectives […] because there are some objectives that are useful intermediaries to the achievement of almost any final goal" (2014: 131f.). Bostrom calls this the instrumental convergence thesis: there are instrumental goals "which are convergent in the sense that their attainment would increase the chances of the agent's goal being realized for a wide range of final goals" (Bostrom 2014: 132).[85] For Bostrom (2014: 132-138), it is likely that a superintelligent AI will have instrumental goals such as self-preservation, cognitive enhancement, technological perfection, and resource acquisition as these help them to achieve whatever their final goals are. Tegmark adds, based on the instrumental convergence principle, a further instrumental goal, namely "truth-seeking and curiosity" (2018: 276). If this is true, it is likely that future AIs will have something like a scientific practice (or whatever the equivalent of "practice" is in the context of the AI form of being). Science (in conjunction with engineering) also allows such an AI to achieve the instrumental goals invoked by Bostrom. For science helps to develop a better world-model which will most likely enhance the system's cognitive capacities (Tegmark 2018: 264), enable it to acquire or create resources more efficiently, and afford better self-preservation strategies.

[85] Similar ideas can already be found in Omohundro (2008).

However, it is more questionable whether AI will have anything like artistic practices as we know them. This is because, first, it is uncertain if future AIs will have phenomenal consciousness. If future AIs will not be phenomenally conscious, they will not have a desire for experiencing aesthetic pleasure and excitement which makes it unlikely that the AI form of being will involve something like artistic practices. Not unrelated to that, future AIs might lack "emotional desires related to […] reproduction" (Tegmark 2018: 311). According to evolutionary aesthetics, the arts are a result of, or at least shaped by, Darwinian selection mechanisms (see, e.g., Dutton 2003; Dutton 2010; Boyd 2005; Miller 2000; Miller 2001). Even if it seems overly reductionistic to explain the origin of art solely in terms of evolutionary processes, it is still questionable if there is a place for artistic practices in the context of an asexual form of being. For these reasons, it is questionable if the AI form of being will feature artistic practices.

However, it is conceptually possible that there will be phenomenally conscious AIs in the future. Yet, the "space of possible minds" (Sloman 1984) is vast. So, it is also conceptually possible that phenomenally conscious AIs "may experience richer, more nuanced mental lives than humans do" (Schneider 2019: 23). To put it differently: "the space of possible AI experiences is huge compared to what we humans can experience […], so we must avoid the pitfall of assuming that being an AI necessarily feels similar to being a person" (Tegmark 2018: 309). Therefore, it is imaginable that future AIs will have something like artistic practices. However, it is unlikely that these will be like the artistic practices of humans. This is not just due to the fact that their mental lives will probably be very different than ours. As AIs can incorporate new modules, they can be "equipped" with new senses: "if we want a computer to 'see' ultraviolet or infrared or 'hear' ultrasound, we can feed the output from an appropriate sensor directly into its 'nervous system[…]'" (Wilczek 2020: 71). Moreover, it is likely that future AIs will live in very different (physical and digital) environments with different

affordances. Thus, if future AIs have something like artistic practices, they will most likely take a radically different form than those that are part of the human form of being.

It needs to be determined whether the concepts introduced in previous chapters make sense when dealing with genuine AI creativity. If they do, the status of value-oriented creativity would become questionable. Given that current ML systems excel at pattern-detection and -imitation, it is likely that the pattern-detecting and -imitating capacities of future AIs will be even more impressive. In many ways, this renders value-oriented creativity much easier for AIs than for humans. Therefore, since value-oriented creativity does not seem to represent a challenge for future AIs, it is not unlikely that the concept of "creativity" will be restricted to sweet-spot and novelty-oriented creativity. At the same time, as mentioned above, the scale and size of the creative efforts of future AIs will most likely exceed those of humans by far. This makes it possible for such machines to run a huge number of large-scale experiments and simulations. These are good preconditions for novelty-oriented creativity – and even sweet-spot creativity, if the system is equipped with the power of judgement (which seems, from a technical perspective, not impossibly difficult to engineer as GANs and CANs show). Furthermore, because of the availability of various (robotic) bodies and the possibility of backing themselves up, future AIs will probably not need to care as much about their physical integrity as humans do. Thus, they can be more daring and bolder than humans when pursuing creative efforts. This might, in many cases, also prove beneficial for novelty-oriented creativity and sweet-spot creativity.

If there will be phenomenally conscious AIs in the future, it is not unlikely that they "might lack our strong human fear of death: as long as they've backed themselves up, all they stand to lose are the memories they've accumulated since their most recent backup" (Tegmark

2018: 311).[86] However, it has been claimed that there is, as Hans Jonas renders it, a "linkage of mortality with creativity in human history" (1992: 39). Many humans are motivated to be creative in order to create something that persists beyond their own finite lives. Does that mean that "immortal" AIs of the future will not be motivated to be creative (unless it helps them to satisfy their basic desires)? This does not necessarily follow. It has been argued that, after a while, immortal conscious beings will suffer from boredom (Williams 1973; May 2009). So, immortal AIs of the future may be motived to be creative to overcome boredom.

Against this, Todd May contends that the arts and other creative efforts will not save immortal beings from getting bored: even if "[g]reat musicians practice for hours a day, day after day", at some point, "it begins to strain credulity to believe that one could stay immersed in a[n artistic] practice for an infinite amount of time" (2009: 61). May focuses exclusively on what we call game-playing creativity, i.e., on participating in a given artistic practice in its current form and perfecting one's ability to do so through training and immersion. This kind of creativity may, indeed, prove inadequate to save an immortal being from becoming bored at some point. However, May's objection disregards other kinds of creativity, i.e., game-changing and game-inventing creativity. Therefore, when participating in existing practices in their current form becomes boring, immortal AIs may change those practices in a radical way so that it remains exciting to participate in them. Or they may establish new practices – and then perfect their ability to participate in them. Thus, creativity – especially game-changing and game-inventing creativity – may enable conscious AIs of the future to overcome the boredom of immortality. In that case, there is a motivation for such machines (if they will ever be built) to be creative, even if they might lack our human fear of death. And there is a special motivation for them to be creative in a game-changing and game-inventing way.

[86] Some may question if it makes sense to speak of "death" with respect to conscious machines. Here, however, we use this terms in a way that does not restrict its application to biological beings. Rather, in this context, "death" shall stand for the irreversible loss of consciousness.

Some of the thoughts presented in this Coda may sound unlikely and fantastical. And maybe they are, in fact, unlikely and fantastical. But the future is not restricted to the next ten years or so. So, maybe, the actual future will be even more unlikely and fantastical than the vision sketched above. It remains to be seen if Paradigm 3 will ultimately emerge and take centre stage. However, if it does not emerge within the next ten years, that does not mean that it will not emerge. The world in which Paradigm 3 prevails will be very different from the worlds Paradigms 1 and 2 emerged in. It will be very different from the world we live in today. What will it be like being human in a world that is also home to "supercreative" posthuman beings? How will Paradigm 3, if it occurs, affect our human self-understanding? The future will show. Maybe human creativity will be powerful enough to find ways of coping with the new situation.

Bibliography

Abrams, M. H. 1972. *The Mirror and the Lamp: Romantic Theory and the Critical Tradition* (Oxford: Oxford University Press)

Addison, Joseph. 1804. *The Works of the Right Honourable Joseph Addison* (Vol. 1) (London: Printed for Vernor and Hood; John Walker; Cuthell and Martin; W.J. and J. Richardson; Longman and Rees; R. Lea; and J. and A. Arch)

Ahmed, Arif. 2010. *Wittgenstein's Philosophical Investigations: A Reader's Guide* (London: Continuum)

Alpaydin, Ethem. 2016. *Machine Learning: The New AI* (Cambridge, MA: MIT Press)

Amabile, Teresa M. 1983. 'Social Psychology of Creativity: A Componential Conceptualization', in *Journal of Personality and Social Psychology* 45(2), pp. 357–76

Amabile, Teresa. 1996. *Creativity in Context* (Boulder, CO: Westview Press)

Amabile, Teresa M. 2013. 'Componential Theory of Creativity', in *Encyclopedia of Management Theory* (Vol. 1), ed. by Eric H. Kessler (Los Angeles, CA: SAGE), pp. 134–39

Andreas-Salomé, Lou. 2001. *Nietzsche*, trans. by Siegfried Mandel (Urbana, IL: University of Illinois Press)

Anscombe, G.E.M. 1957. *Intention* (Oxford: Basil Blackwell)

Aquinas, Thomas. 1891. *Summa Theologica: Diligenter Emendata: De Rubeis, Billuart et Aliorum: Notis Selectis Ornata* (Vol. 1ª), ed. by Petrus Marietti (Augusta Taurinorum / Turin: Typographia Pontificia et Archiepiscopalis)

Aristotle. 2013. *Poetics*, trans. by Anthony Kenny (Oxford: Oxford University Press)

Aristotle. 2018. *Physics*, trans. by C. D. C. Reeve (Indianapolis, IN; Cambridge, MA: Hackett Publishing Company, Inc.)

Asma, Stephen T. 2017. *The Evolution of Imagination* (Chicago, IL; London: University of Chicago Press)

Attneave, Fred. I960. 'In Defence of Homunculi', in *Sensory Communication*, ed. by Walter A. Rosenblith (Cambridge, MA: MIT Press), pp. 777–82

Audi, Robert. 2018. 'Creativity, Imagination, and Intellectual Virtue', in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 25–41

Awodey, Steve. 2006. *Category Theory* (Oxford: Oxford University Press)

Bacon, Francis. 1964. 'The Refutation of Philosophies', in *The Philosophy of Francis Bacon*, ed. by Benjamin Farrington (Chicago, IL: University of Chicago), pp. 103–32

Baker, Gordon P. 2004. *Wittgenstein's Method: Neglected Aspects: Essays on Wittgenstein*, ed. by Katherine J. Morris (Oxford: Blackwell)

Baker, Gordon P., and P.M.S. Hacker. 2005. *Wittgenstein: Understanding and Meaning: Volume 1 of An Analytical Commentary on the Philosophical Investigations* (Oxford: Blackwell)

Baker, Mona. 2006. *Translation and Conflict: A Narrative Account* (London: Routledge)

Barthes, Roland. 1986. 'The Death of the Author', in *The Rustle of Language*, trans. by Richard Howard (New York, NY: Hill and Wang), pp. 49–55

Baudrillard, Jean. 2005. *The Conspiracy of Art: Manifestos, Interviews, Essays* (Cambridge, MA; London: Semiotext)

Batteux, Charles. 2016. *The Fine Arts Reduced to a Single Principle* (Oxford: Oxford University Press)

Bayne, Tim, 2017. 'Free Will and the Phenomenology of Agency', in *The Routledge Companion to Free Will*, ed. by Kevin Timpe, Meghan Griffith, and Neil Levy (New York, NY: Routledge), 633–44

Beaney, Michael. 2005. *Imagination and Creativity* (Milton Keynes: Open University Press)

Beja, Morris. 1992. *James Joyce: A Literary Life* (Basingstoke: Macmillan)

Ben-Yami, Hanoch. 2017. 'Vagueness and Family Resemblance', in *A Companion to Wittgenstein*, ed. by Hans-Johann Glock and John Hyman (Hoboken: Wiley), pp. 407–19

Blair, Hugh. 1823. *Lectures on Rhetoric and Belles Lettres* (London: Printed for Baynes and Son, and Walker and Co.)

Block, Ned. 1995. 'On a Confusion about a Function of Consciousness', in *The Behavioral and Brain Sciences* 18(2), pp. 227–47

Blumenberg, Hans. 2000. '"Imitation of Nature": Toward a Prehistory of the Idea of the Creative Being', in *Qui parle* 12 (1), pp. 17–54

Boden, Margaret A. 1970. 'Intentionality and Physical Systems', in *Philosophy of Science* 37(2), pp. 200–14

Boden, Margaret. 2004. *The Creative Mind: Myths and Mechanisms* (London: Routledge)

Boden, Margaret. 2009a. 'Creativity: How does it Work?', in *The Idea of Creativity*, ed. by Michael Krausz, Denis Dutton, and Karen Bardsley (Leiden; Boston: Brill), pp. 237–50

Boden, Margaret. 2009b. 'Conceptual Spaces', in *Milieus of Creativity: An Interdisciplinary Approach to Spatiality of Creativity* (Vol. 2), ed. by Peter Meusburger, Joachim Funke, and Edgar Wunder (Dordrecht: Springer), pp. 235–43

Boden, Margaret A. 2014. 'Creativity and Artificial Intelligence: A Contradiction in Terms?', in *The Philosophy of Creativity: New Essays*, ed. by Elliot Samuel Paul and Scott Barry Kaufman (New York, NY: Oxford University Press), pp. 224–44

Boghossian, Paul A. 1989. 'The Rule-Following Considerations', in *Mind* 98(392), pp. 507–49

Boghossian, Paul A. 2015. 'Is (Determinate) Meaning a Naturalistic Phenomenon?', in *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism*,

ed. by Steven Gross, Nicholas Tebben, and Michael Williams (Oxford: Oxford University Press), pp. 331–58

Bostrom, Nick. 2005. 'A History of Transhumanist Thought', in *Journal of Evolution and Technology* 14(1), pp. 1–25

Bostrom, Nick. 2008. 'Why I Want to Be a Posthuman When I Grow Up', in *Medical Enhancement and Posthumanity*, ed. by Bert Gordijn and Ruth Chadwick (Dordrecht: Springer), pp. 107–36

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press)

Boyd, Brian. 2005. 'Evolutionary Theories of Art', in *The Literary Animal: Evolution and the Nature of Narrative*, ed. by Jonathan Gottschall and David Sloan Wilson (Evanston, IL: Northwestern University Press), pp. 147–76

Brand, Myles. 1984. *Intending and Acting: Toward a Naturalized Action Theory* (Cambridge, MA: MIT Press)

Brandom, Robert. 1994. *Making it Explicit* (Cambridge, MA: Harvard University Press)

Brassaï. 1999. *Conversations with Picasso*, ed. by Jane Marie Todd (Chicago, IL; London: University of Chicago Press)

Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press)

Braver, Lee. 2012. *Groundless Grounds: A Study of Wittgenstein and Heidegger* (Cambridge, MA: MIT Press)

Brooks, Michael. 2019. 'Beyond Quantum Supremacy: The Hunt for Useful Quantum Computers', in *Nature* 574(7776), pp. 19–21

Brooks, Rodney Allen. 2002. *Flesh and Machines: How Robots Will Change Us* (London: Allen Lane)

Brown, Glen R. 1999. 'Postmodernism and Creativity', in *Encyclopedia of Creativity* (Vol. 2), ed. by Mark A. Runco and Steven R. Pritzker, pp. 423–28

Brown, Jonathan. 1996. 'Picasso and the Spanish Tradition of Painting', in *Picasso and the Spanish Tradition*, ed. by Jonathan Brown (New Haven, CT: Yale University Press), pp. 1–25

Buonarroti, Michelangelo. 1991. *The Poetry of Michelangelo*, trans. by James S. Saslow (New Haven, CN; London: Yale University Press)

Butler, Samuel. 1923. 'Darwin Among the Machines', in *A First Year in Canterbury Settlement and Other Early Essays* (London: J. Cape; New York, NY: E. P. Dutton), pp. 208–13

Byrne, David. 2012. *How Music Works* (Edinburgh: Canongate)

Campbell, Donald T. 1960. 'Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes', in *Psychological Review* 67(6), pp. 380–400

Carroll, Noël. 2014. 'The Creative Audience: Some Ways in which Readers, Viewers, and/or Listeners Use Their Imaginations to Engage Fictional Artworks', in *The Philosophy of*

Creativity: New Essays, ed. by Elliot Samuel Paul and Scott Barry Kaufman (New York, NY: Oxford University Press), pp. 62–81

Carruthers, Peter. 2006. *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought* (Oxford: Clarendon Press)

Carruthers, Peter, and Elizabeth Picciuto. 2014. 'The Origins of Creativity', in *The Philosophy of Creativity*, ed. by Elliot Samuel Paul and Scott Barry Kaufman (Oxford: Oxford University Press), pp. 199–223

Cassirer, Ernst. 2009. *The Philosophy of the Enlightenment* (Princeton,NJ; Oxford: Princeton University Press)

Cavell, Stanley. 1979. *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy* (Oxford: Oxford University Press)

Cavell, Stanley. 1988. 'Declining Decline: Wittgenstein as a Philosopher of Culture', in *Inquiry* 31(3), pp. 253–64

Cavell, Stanley. 2013. *This New Yet Unapproachable America: Lectures after Emerson after Wittgenstein* (Chicago, IL: University of Chicago Press)

Cavell, Stanley. 2018. *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism: The Carus Lectures, 1988* (Chicago, IL: The University of Chicago Press)

Caws, Mary Ann. 2005. *Pablo Picasso* (London: Reaktion Books)

Chalmers, David J. 2016. 'The Singularity: A Philosophical Analysis', in *The Singularity: Could Artificial Intelligence Really Out-Think Us (and Would We Want It To)?*, ed. by Uziel Awret (Exeter: Imprint Academic), pp. 517–48

Child, William. 2011. *Wittgenstein* (London: Routledge)

Child, William. 2019. 'Meaning, Use, and Supervenience', in *Wittgenstein on Philosophy, Objectivity, and Meaning*, ed. by James Conant and Sebastian Sunday (Cambridge: Cambridge University Press), pp. 211–30

Child, William. 2020. '"We Can Go No Further": Meaning, Use, and the Limits of Language', in *Wittgenstein and the Limits of Language*, ed. by Hanne Appelqvist (New York, NY: Routledge), pp. 93–114

Choi, Yejin. 2022. 'The Curious Case of Commonsense Intelligence', in *Daedalus* 151(2), pp. 139–55

Cicero, Marcus Tullius. 1892. 'On the Nature of the Gods', in *The Treatises of M.T. Cicero: On the Nature of the Gods, On Divination, On Fate, On the Republic, On the Laws, and On Standing for the Consulship*, trans. by Charles Duke Yonge (London; New York, NY: George Bell & Sons), pp. 1–140

Clark, Andy. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford: Oxford University Press)

Clark, Andy, and David Chalmers. 1998. 'The Extended Mind', in *Analysis* 58(1), pp.7–19

Clynes, Manfred E., and Nathan S. Kline. 1960. ‘Cyborgs and Space’, in *Astronautics* 5(9), pp. 26–27, 74–76

Coeckelbergh, Mark, and Michael Funk. 2018. ‘Wittgenstein as a Philosopher of Technology: Tool Use, Forms of Life, Technique, and a Transcendental Argument’, in *Human Studies: A Journal for Philosophy and the Social Sciences* 41(2), pp. 165–91

Cohn, Gabe. 2018. ‘AI Art at Christie’s Sells for $432,500’, in *The New York Times*, 25 October 2018 (https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html; accessed 21/11/2022)

Conant, James. 1998. ‘Wittgenstein on Meaning and Use’, in *Philosophical Investigations* 21(3), pp. 222–50

Conant, James. 2004. ‘A Ladder Turns into a Fly-Bottle’, in *New History of German Literature*, ed. by David Wellbery, Hans Ulrich Gumbrecht, Anton Kaes, Dorothea von Muecke, and Judith Ryan (Cambridge, MA: Harvard University Press), pp. 866–71

Corry, Leo. 2004. *Modern Algebra and the Rise of Mathematical Structures* (Basel: Springer Basel)

Crane, Tim. 2015. ‘The Mental States of Persons and their Brains’, in *Royal Institute of Philosophy Supplement* 76, pp. 253–70

Cropley, David H. 2019. *Homo Problematis Solvendis–Problem-Solving Man: A History of Human Creativity* (Singapore: Springer)

Cropley, David H., James C. Kaufman, and Arthur J. Cropley. 2008. ‘Malevolent Creativity: A Functional Model of Creativity in Terrorism and Crime’, in *Creativity Research Journal* 20(2), pp. 105–15

Crutchfield, Richard S. 1962. ‘Conformity and Creative Thinking’, in *Contemporary Approaches to Creative Thinking*, ed. Howard E. Gruber, Glenn Terrell, and Michael Wertheimer (New York, NY: Atherton Press), pp. 120–40

Csíkszentmihályi, Mihály. 1999. ‘Implications of a Systems Perspective for the Study of Creativity’, in *Handbook of Creativity*, ed. by Robert J. Sternberg (Cambridge: Cambridge University Press), pp. 313–36

Csíkszentmihályi, Mihály. 2013. *Creativity: The Psychology of Discovery and Invention* (New York City, NY: Harper Perennial)

Curtius, Ernst Robert. 2013. *European Literature and the Latin Middle Ages*, trans. by Willard R. Trask (Princeton, NJ: Princeton University Press)

Cusa, Nicholas of. 1998. ‘De Beryllo (On [Intellectual] Eyeglasses)’, in *Nicholas of Cusa: Metaphysical Speculations*, trans. by Jasper Hopkins (Minneapolis, MN: Banning Press), pp. 791–838 (https://jasper-hopkins.info/DeBeryllo12-2000.pdf; accessed 13/11/2021)

Danto, Arthur C. 1981. *The Transfiguration of the Commonplace: A Philosophy of Art* (Cambridge, MA: Harvard University Press)

Dartnall, Terry. 1994. ‘Creativity, Combination, and Cognition’, in *Behavioral and Brain Sciences* 17(3), p. 537

Darwin, Charles. 2008. *The Origin of Species*, ed. by Gillian Beer (Oxford: Oxford University Press)

Davidson, Donald. 1963. 'Actions, Reasons, and Causes', in *The Journal of Philosophy* 60(23), pp. 685–700

Descartes, René. 2017. *Meditations on First Philosophy: With Selections from the Objections and Replies*, ed. by John Cottingham (Cambridge: Cambridge University Press)

Dennett, Daniel C. 1969. *Content and Consciousness* (London: Routledge & Kegan Paul)

Dennett, Daniel C. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, MA: MIT Press; Bradford Books)

Dennett, Daniel. 2004. 'Can Machines Think?', in *Alan Turing: Life and Legacy of a Great Thinker*, ed. by Christof Teuscher (Berlin: Springer), pp. 295–316

Dennett, Daniel C. 2014. *Intuition Pumps and Other Tools for Thinking* (London: Penguin Books)

Dennett, Daniel C. 2017. *From Bacteria to Bach and Back: The Evolution of Minds* (New York, NY: W.W. Norton & Company)

Diamond, Cora. 1989. 'Rules: Looking in the Right Place', in *Wittgenstein: Attention to Particulars: Essays in Honour of Rush Rhees*, ed. by D. Z. Phillips and Peter Winch (Basingstoke: Macmillan), pp. 12–34

Douglas, Heather E. 2016. 'Values in Science', in *The Oxford Handbook of Philosophy of Science*, ed. by Paul Humphreys, pp. 609–30

Dresler, Martin. 2008. 'Einleitung: Kreativität als Offenes Konzept', in *Kreativität: Beiträge aus den Natur- und Geisteswissenschaften*, ed. by Martin Dresler and Tanja Gabriele Baudson (Stuttgart: Hirzel), pp. 5–20

Dreyfus, Hubert L. 1972. *What Computers Can't Do: A Critique of Artificial Reason* (New York, NY; London: Harper & Row)

Duff, William. 1767. *An Essay on Original Genius; and its Various Modes of Exertion in Philosophy and the Fine Arts, Particularly in Poetry* (London: Printed for Edward and Charles Dilly in the Poultry, near the Mansion-House)

Dummett, Michael. 1978. *Truth and other Enigmas* (London: Duckworth)

du Sautoy, Marcus. 2019. *The Creativity Code: How AI is Learning to Write, Paint and Think* (London: Fourth Estate)

Dutton, Denis. 2003. 'Aesthetics and Evolutionary Psychology', in *The Oxford Handbook for Aesthetics*, ed. by Jerrold Levinson (Oxford: Oxford University Press), pp. 693–705

Dutton, Denis. 2010. *The Art Instinct: Beauty, Pleasure, and Human Evolution* (Oxford: Oxford University Press)

Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. 'CAN: Creative Adversarial Networks, Generating "Art" by Learning about Styles and Deviating from Style Norms' (arXiv:1706.07068v1; accessed 30/10/2019)

Erden, Yasemin J. 2010. ‘Could a Created Being Ever be Creative? Some Philosophical Remarks on Creativity and AI Development’, in *Minds & Machines* 20(3), pp. 349–62

Eriksen, Cecilie. 2020. ‘Winds of Change: The Later Wittgenstein’s Conception of the Dynamics of Change’, in *Nordic Wittgenstein Review* 9, pp. 79–105

Fauconnier, Gilles. 2010. ‘Mental Spaces’, in *The Oxford Handbook of Cognitive Linguistics*, ed. by Dirk Geeraerts and Herbert Cuyckens (New York, NY: Oxford University Press), pp. 351–76

Fauconnier, Gilles, and Mark Turner. 1995. ‘Conceptual Integration and Formal Expression’, in *Metaphor and Symbol* 10(3), pp. 183–204

Fauconnier, Gilles, and Mark Turner. 1998. ‘Conceptual Integration Networks’, in *Cognitive Science* 22(2), pp. 133–87

Fauconnier, Gilles, and Mark Turner. 2002. *The Way We Think. Conceptual Blending and The Mind's Hidden Complexities* (New York, NY: Basic Books)

Fauconnier, Gilles, and Mark Turner. 2003. ‘Conceptual Blending, Form and Meaning’, in *Recherches en Communication* 19(19), pp. 57–86

Feynman, Richard P. 1982. ‘Simulating Physics with Computers’, in *International Journal of Theoretical Physics* 21(6-7), pp. 467–88

Finke, Ronald A., Steven M. Smith, Thomas B. Ward. 1992. *Creative Cognition: Theory, Research and Applications* (Cambridge, MA; London: MIT Press)

Finke, Ronald A., Steven M. Smith, Thomas B. Ward. 1999. ‘Creative Cognition’, in *Handbook of Creativity*, ed. by Robert J. Sternberg (Cambridge: Cambridge University Press), pp. 189–212

Fizeau, M. H. 1849. ‘Sur une expérience relative à la vitesse de propagation de la lumière’, in *Comptes Rendus de l’Académie des Sciences* 29, pp. 90–92

Floridi, Luciano, and J.W. Sanders. 2004. ‘On the Morality of Artificial Agents’, in *Minds and Machine* 14, pp. 349–79

Floyd, Juliet. 2017. ‘Turing on “Common Sense”: Cambridge Resonances’, in *Philosophical Explorations of the Legacy of Alan Turing: Turing 100* (Cham: Springer), pp. 103–52

Fodor, Jerry A. 1968. *Psychological Explanation: An Introduction to the Philosophy of Psychology* (New York: Random House)

Forster, Michael. 2010. ‘Wittgenstein on Family Resemblance Concepts’, in *Wittgenstein’s Philosophical Investigations: A Critical Guide*, ed. by Arif Ahmed (Cambridge: Cambridge University Press), pp. 66–87

Foucault, J. L. 1862. ‘Détermination expérimentale de la vitesse de la lumière; parallaxe du Soleil’, in *Comptes Rendus de l’Académie des Sciences* 55, pp. 501–03, 792–96

Frege, Gottlob. 1956. ‘The Thought: A Logical Inquiry’, in *Mind* 65(259), pp. 289–311

Frege, Gottlob. 1960. *Translations from the Writings of Gottlob Frege*, ed. and trans. by Peter Geach and Max Black (Oxford: Blackwell)

Galton, Francis. 1883. *Inquiries into Human Faculty and its Development* (London: Macmillan)

Galton, Francis. 1892. *Hereditary Genius: An Inquiry into its Laws and Consequences* (London: Macmillan)

Gangadharbatla, Harshavardhan. 2010. 'Technology Component: A Modified Systems Approach to Creative Thought', in *Creativity Research Journal*, 22(2), pp. 219–27

Gardner, Howard. 1993. *Creating Minds: An Anatomy of Creativity Seen Through the Lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham, and Gandhi* (New York, NY: BasicBooks)

Gaut, Berys. 2003. 'Creativity and Imagination', in *The Creation of Art: New Essays in Philosophical Aesthetics*, ed. by Berys Gaut and Paisley Livingstone (Cambridge: Cambridge University Press), pp. 148–73

Gaut, Berys. 2010. 'The Philosophy of Creativity', in *Philosophy Compass* 5(12), pp. 1034–46

Gaut, Berys. 2018. 'The Value of Creativity', in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 124–39

Gatys, Leon A., Alexander S. Ecker, Matthias Bethge. 2015. 'A Neural Algorithm of Artistic Style' (https://arxiv.org/abs/1508.06576; accessed 09/05/2019)

Gerard, Alexander. 1774. *An Essay on Genius* (London: Printed for W. Strahan, T. Cadell, and W. Creech)

Gibson, James J. 2015. *The Ecological Approach to Visual Perception* (New York, NY; Hove: Psychology Press)

Ginet, Carl. 1990. *On Action* (Cambridge: Cambridge University Press)

Glăveanu, Vlad P., and James C. Kaufman. 2019. 'Creativity: A Historical Perspective', in *The Cambridge Handbook of* Creativity, ed. by James C. Kaufman and Robert J. Sternberg (Cambridge: Cambridge University Pres), pp. 9–26

Glock, Hans-Johann. 1996. *A Wittgenstein Dictionary* (Oxford: Blackwell)

Godfrey-Smith, Peter. 2009. *Darwinian Populations and Natural Selection* (Oxford: Oxford University Press)

Goertzel, Ben. 2019. 'SingularityNET CEO Discusses the Future of AI' (https://itbrief.com.au /story/singularitynet-ceo-discusses-the-future-of-ai; accessed 27/03/2022)

Goethe, Johann Wolfgang. 1985. 'Speech on the Shakespeare Day', trans. by David Paisey, in *Comparative Criticism* 7, pp. 177–81

Gombrich, E. H. 1962. *Art and Illusion: A Study in the Psychology of Pictorial Representation* (London: Phaidon Press)

Good, I. J. 1965. 'Speculations Concerning the First Ultraintelligent Machine', in *Advanced in Computers* 6, pp. 31–88

Grant, James. 2012. 'The Value of Imaginativeness', in *Australasian Journal of Philosophy* 90(2), pp. 275–89

Grant, Rhiannon, and Myfanwy Reynolds. 2016. ‘Can Chewie Speak? Wittgenstein and the Philosophy of Language’, in *The Ultimate Star Wars and Philosophy: You Must Unlearn What You Have Learned*, ed. by Jason T. Eberl and Kevin S. Decker (Chichester, West Sussex: Wiley Blackwell), pp. 240–49

Getzels, Jacob W. 1964. ‘Creative Thinking, Problem-Solving, and Instruction’, in *Theories of Learning and Instruction*, ed. by Ernest R. Hilgard (Chicago: Chicago University Press), pp. 240–67

Goldfarb, Warren. 1985. ‘Kripke on Wittgenstein on Rules’, in *The Journal of Philosophy* 82(9), pp. 471–88

Grayling, A. C. 2001. *Wittgenstein: A Very Short Introduction* (Oxford: Oxford University Press)

Grève, Sebastian Sunday, and Jakub Mácha. 2016. ‘The Good, the Bad and the Creative: Language in Wittgenstein’s Philosophy’, in *Wittgenstein and the Creativity of Language* (Basingstoke, Hampshire: Palgrave Macmillan), pp. 3–25

Grice, H. Paul. 1975. ‘Logic and Conversation’, in *Syntax and Semantics 3: Speech Acts*, ed. by Peter Cole and Jerry L. Morgan (New York, NY: Academic Press), pp. 41–58

Grice, H. Paul. 1989. *Studies in the Way of Words* (Cambridge, MA; London: Harvard University Press)

Guilford, J. P. 1950. Creativity, in *American Psychologist* 5(9), pp. 444–54

Guilford, J. P. 1956. ‘The Structure of Intellect’, in *Psychological Bulletin* 53(4), pp. 267–93

Guyer, Paul. 2003. ‘Exemplary Originality: Genius, Universality, and Individuality’, in *The Creation of Art*, ed. by Berys Gaut and Paisley Livingston (New York, NY: Cambridge University Press), pp. 116–37

Haase, Matthias. 2012. ‘Three Forms of the First Person Plural’, in *Rethinking Epistemology* (Vol. 2.), ed. by Günter Abel and James Conant (Berlin; Boston, MA: De Gruyter), pp. 229–56

Hacker, P.M.S. 1997. *Wittgenstein: On Human Nature* (London: Phoenix)

Hacker, P.M.S. 2007. ‘Gordon Baker’s Late Interpretation of Wittgenstein’, in *Wittgenstein and his Interpreters: Essays in Memory of Gordon Baker*, ed. by Guy Kahane, Edward Kanterian, and Oskari Kuusela (Oxford: Blackwell)

Hacker, P.M.S. 2013. *Wittgenstein: Comparisons and Context* (Oxford: Oxford University Press)

Hacker, P. M. S. 2019. *Wittgenstein: Meaning and Mind: An Analytical Commentary on the Philosophical Investigations* (Vol. 3) (Hoboken, NJ: Wiley Blackwell)

Hacking, Ian. 2012. ‘Introductory Essay’, in Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, IL; London: University of Chicago Press), pp. vii–xxxvii

Hadamard, Jacques. 1945. *An Essay on the Psychology of Invention in the Mathematical Field* (Princeton, NJ: Princeton University Press)

Hadjeres, Gaëtan, François Pache, and Frank Nielsen. 2017. ‘DeepBach: A Steerable Model for Bach Chorales Generation’ (arXiv:1612.01010; accessed 09/05/2019)

Halper, Edward. 1989. ‘Is Creativity Good?’, in *British Journal of Aesthetics* 29(1), pp. 47–56

Harari, Yuval N. 2015. *Homo Deus: A Brief History of Tomorrow* (London: Harvill Secker)

Hartree, Douglas Rayner. 1949. *Calculating Instruments and Machines* (University of Illinois Press: Urbana)

Haugeland, John. 1985. *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT Press)

Hebb, Donald O. 1980. *Essay on Mind* (Hillsdale, NJ: Lawrence Erlbaum)

Helmholtz, Hermann von. 1896. ‘Erinnerungen’ (1891), in *Vorträge und Reden* (Vol. 1) (Braunschweig: Friedrich Vieweg und Sohn), pp. 1–22

Herder, Johann Gottfried. 2008. *Shakespeare*, ed. and trans. by Gregory Moore (Princeton, NJ: Princeton University Press)

Hills, Alison, and Alexander Bird. 2018. ‘Creativity without Value’, in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 95–107

Hills, Alison, and Alexander Bird. 2019. ‘Against Creativity’, in *Philosophy and Phenomenological Research* 99(3), pp. 694–713

Höffe, Otfried. 2001. *Categorical Principles of Law: A Counterpoint to Modernity*, trans. by Mark Migotti (University Park, PA: Pennsylvania State University Press)

Hofstadter, Douglas R. 1980. *Gödel, Escher, Bach: An Eternal Golden Braid* (Harmondsworth: Penguin)

Holton, Richard. 2001. ‘Meaning and Rule-following: Philosophical Aspects’, in *International Encyclopedia of the Social & Behavioral Sciences*, ed. by Neil J. Smelser and Paul B. Baltes, pp. 9431–35

Homer. 1996. *The Odyssey*, trans. by Robert Fagles (New York; London: Penguin)

Horkheimer, Max, and Theodor W. Adorno. 2002. *Dialectic of Enlightenment: Philosophical Fragments*, ed. by Gunzelin Schmid Noerr, trans. by Edmund Jephcott (Stanford, CA: Stanford University Press)

Horwich, Paul. 2012. *Wittgenstein’s Metaphilosophy* (Oxford: Oxford University Press)

Hume, David. 1810. *The History of England, From the Invasion of Julius Cæsar to the Revolution in 1688* (Pennsylvania, PA: Published by Levis & Weaver)

Hume, David. 1993. ‘Of the Rise and Progress of the Arts and Sciences’, in *Selected Essays*, ed. by Stephen Copley and Andrew Edgar (Oxford; New York, NY: Oxford University Press), pp. 56–76

Hutchinson, Phil, and Rupert Read. 2008. ‘Toward a Perspicuous Presentation of “Perspicuous Presentation”’, in *Philosophical Investigations* 31(2), pp. 141–60

Huxley, Julian. 1957. ‘Transhumanism’, in *New Bottles for New Wine: Essays* (London: Chatto & Windus), pp. 13–17

Hsü, Kenneth J. 1986. ‘Darwin’s Three Mistakes’, in *Geology* 14(6), pp. 532–34

Johnson, Samuel. 1846. *The Works of Samuel Johnson, LL.D: With an Essay on his Life and Genius by Arthur Murphy* (Vol. 1) (New York, NY: Alexander V. Blake Publisher)

Jonas, Hans. 1992. 'The Burden and Blessing of Mortality', in *The Hastings Center Report* 22(1), pp. 34–40

Jordanous, Anna, and Bill Keller. 'Modelling Creativity: Identifying Key Components through a Corpus-Based Approach', in *PLoS ONE* 11(10), pp. 1–27

Kadanoff, Leo P. 2013. 'Theories of Matter: Infinities and Renormalization', in *The Oxford Handbook of Philosophy of Physics*, ed. by Robert Batterman (Oxford: Oxford University Press), pp. 141–88

Kakar, Sudhir, and Günter Blamberger (ed). 2015. *On Creativity* (New York, NY: Penguin Viking)

Kane, Robert. 1999. *The Significance of Free Will* (New York, NY: Oxford University Press)

Kaufman, James C., and Robert J. Sternberg. 2007. 'Creativity', in *Change: The Magazine of Higher Learning* 39(4), pp. 55–60

Kaufman, James C., and Vlad Glăveanu. 2019. 'A Review of Creativity Theories: What Questions Are We Trying to Answer?', in *The Cambridge Handbook of Creativity*, ed. by James C. Kaufman & Robert J. Sternberg (Cambridge: Cambridge University Press), pp. 27–43

Kant, Immanuel. 2000. *Critique of the Power of Judgment*, ed. by Paul Guyer, trans. by Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press)

Kant, Immanuel. 2012. *Groundwork of the Metaphysics of Morals*, ed. by Jens Timmermann and Mary Gregor (Cambridge: Cambridge University Press)

Kasparov, Garry, and Mig Greengard. 2018. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins* (London: John Murray)

Kelly, Sean Dorrance. 2019. 'What Computers can't Create', in *MIT Technology Review* 122, pp. 68–75

Kieran, Matthew. 2014. 'Creativity as a Virtue of Character', in *The Philosophy of Creativity: New Essays*, ed. by Elliot Samuel Paul and Scott Barry Kaufman (New York, NY: Oxford University Press), pp. 125–44

Kieran, Matthew. 2018. 'Creativity, Vanity and Narcissism', in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 74–92

Kind, Amy. 2022. *Imagination and Creative Thinking* (Cambridge: Cambridge University Press)

Kivy, Peter. 2013. 'Genius and the Creative Imagination', in *Oxford Handbook of British Philosophy in the Eighteenth Century*, ed. by James A. Harris (Oxford: Oxford University Press), pp. 468–88

Koestler, Arthur. 1964. *The Act of Creation* (New York, NY: Penguin Books)

Koestler, Arthur. 2009. 'The Three Domains of Creativity', in *The Idea of Creativity*, ed. by Michael Krausz, Denis Dutton, and Karen Bardsley (Leiden; Boston, MA: Brill), pp. 251–66

Kozbelt, Aaron. 2011. ‘Theories of Creativity’, in *Encyclopedia of Creativity*, ed. by Mark A. Runco and Steven R. Pritzker (Amsterdam; Boston, MA: Academic Press/Elsevier), pp. 473–79

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition* (Cambridge, MA: Harvard University Press)

Kristeller, Paul Oskar. 2021. ‘The Modern System of the Arts’, in *Renaissance Thought and the Arts: Collected Essays* (Princeton, NJ: Princeton University Press), pp. 163–227

Kronfeldner, Maria. 2009. ‘Creativity Naturalized’, in *The Philosophical Quarterly* 59(237), pp. 577–92

Kronfeldner, Maria. 2018. ‘Explaining Creativity’, in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 213–29

Kuhn, Thomas S. 1978a. ‘Objectivity, Value Judgment, and Theory Choice’, in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (University of Chicago Press), pp. 320–39

Kuhn, Thomas S. 1978b. ‘Comment on the Relations of Science and Art’, in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (University of Chicago Press), pp. 340–51

Kuhn, Thomas S. 1991. ‘The Road since Structure’, in *PSA 1990: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 2), ed. by Linda Wessels, Arthur Fine, and Micky Forbes (East Lansing, MI: Philosophy of Science Association), pp. 3–13

Kuhn, Thomas S. 2012. *The Structure of Scientific Revolutions* (Chicago, IL; London: University of Chicago Press)

Kullmann, Wolfgang. 2010. ‘The Contemplative and The Political Forms of Life (Book X.6–9)’, in *Aristotle’s “Nicomachean Ethics”*, ed. by Otfried Höffe (Leiden, the Netherlands; Boston: Brill), pp. 205–23

Kurzweil, Ray. 2004. ‘The Law of Accelerating Returns’, in *Alan Turing: Life and Legacy of a Great Thinker*, ed. by Christof Teuscher (Berlin: Springer), pp. 381–416

Kurzweil, Ray. 2010. *The Singularity is Near* (London: Gerald Duckworth & Co)

Lakatos, Imre. 2004. ‘Falsification and the Methodology of Scientific Research Programmes’, in *Criticism and the Growth of Knowledge: International Colloquium in the Philosophy of Science (1965: Bedford College)*, ed. by Imre Lakatos and Alan Musgrave (Cambridge: Cambridge University Press), pp. 91–196

Langland-Hassan, Peter. 2020. *Explaining Imagination* (Oxford: Oxford University Press)

Larson, Erik J. 2021. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do* (Cambridge, MA; London: The Belknap Press of Harvard University Press)

Lavater, Johann Caspar. 1790. *Aphorisms on Man: Translated from the Original Manuscript of the Rev. John Caspar Lavater, Citizen of Zuric* (Dublin: W. Sleater, Dame-Street, and P. Byrne, Grafton-Street)

Leach, Neil. 2022. ‘In the Mirror of AI: What is Creativity?’, in *Architectural Intelligence* 1(15), pp. 1–8

Lee, Kai-Fu, and Chen Qiufan. 2021. *AI 2041: Ten Visions for Our Future* (London: W H Allen)

Lehrer, Keith. 1960. ‘Can We Know That We Have Free Will by Introspection?’, in *The Journal of Philosophy* 57(5), pp. 145–57

Lendvai, Ernő. 1979. *Bela Bartók: An Analysis of his Music* (London: Kahn & Averill)

Levi, Isaac. 1960. ‘Must the Scientist Make Value Judgments?’, in *The Journal of Philosophy* 57(11), pp. 345–57

Lewis, David K. 1986. *The Plurality of Worlds* (Oxford: Basil Blackwell)

Liu, Lydia H. 2021. ‘Wittgenstein in the Machine’, in *Critical Inquiry* 47(3), pp. 425–55

Livingston, Paisley. 2018. ‘Explicating “Creativity”’, in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 108–23

Lloyd, Seth. 2000. ‘Ultimate Physical Limits to Computation’, in *Nature* 406(6799), pp. 1047–54

Locke, John. 1975. *An Essay Concerning Human Understanding*, ed. by Peter H. Nidditch (Oxford: Clarendon Press)

Loos, Adolf. 2012. *Ornament und Verbrechen*, ed. by Peter Stuiber (Vienna: Metroverlag)

Longinus. 1890. *On the Sublime*, trans. by H. L. Havell (London: Macmillan)

Loran, Erle. 2006. *Cézanne’s Composition: Analysis of his Form with Diagrams and Photographs of his Motifs* (Berkeley, CA: University of California Press)

Lowinsky, Edward E. 1964. ‘Musical Genius: Evolution and Origins of a Concept—II’, in *The Musical Quarterly* 50(4), pp. 476–95

Lucas, John R. 1961. ‘Minds, Machines and Gödel’, in *Philosophy* 36(137), pp. 112–27

Lucretius Carus, Titus. 2007. *The Nature of Things*, trans. by A.E. Stallings (London: Penguin)

Luper, Steven. 2019. ‘Death’, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta (https://plato.stanford.edu/archives/win2019/entries/death/; accessed 01/11/2020)

Macarthur, David. 2014. ‘Cavell on Skepticism and the Importance of Not-Knowing’, in *Conversations: The Journal of Cavellian Studies* 2, pp. 2–23

Mackenzie, Tom. 2021. *Poetry and Poetics in the Presocratic Philosophers: Reading Xenophanes, Parmenides and Empedocles as Literature* (Cambridge: Cambridge University Press)

MacLeod, Christine. 2009. ‘The Invention of Heroes’, in *Nature* 460(7255), pp. 572–73

Malcolm, Norman. 1986. *Nothing is Hidden* (Oxford: Blackwell)

Malthus, Thomas Robert. 2008. *An Essay on the Principle of Population*, ed. by Geoffrey Gilbert (Oxford: Oxford University Press)

Markel, Howard. 2020. *Literatim: Essays at the Intersections of Medicine and Culture* (New York, NY: Oxford University Press)

Martin, Mike W. 2007. *Creativity: Ethics and Excellence in Science* (Lanham, MD; Plymouth: Lexington Books)

Maurice, Donald. 2004. *Bartók's Viola Concerto: The Remarkable Story of His Swansong* (New York, NY: Oxford University Press)

Mason, John Hope. 1988. 'The Character of Creativity: Two Traditions', in *History of European Ideas* 9(6), p. 697–715

Mason, John Hope. 2003. *The Value of Creativity: The Origins and Emergence of a Modern Belief* (Aldershot: Ashgate)

May, Todd. 2009. *Death* (Stocksfield: Acumen)

McCloskey, Michael, and Neal J. Cohen. 1989. 'Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem', in *Psychology of Learning and Motivation* 24, pp. 109–65

McCorduck, Pamela. 1979. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (San Francisco: W.H. Freeman)

McDowell, John. 1984. 'Wittgenstein on Following a Rule', in *Synthese* 58(3), pp. 325–63

McDowell, John. 1992. 'Meaning and Intentionality in Wittgenstein's Later Philosophy', in *Midwest Studies in Philosophy* 17(1), pp. 40–52

McDowell, John. 2009. 'Wittgensteinian "Quietism"', in *Common Knowledge* 15(3), pp. 365–72

McGinn, Marie. 2013. *The Routledge Guidebook to Wittgenstein's Philosophical Investigations* (Abingdon; New York, NY: Routledge)

McLaren, Robert B. 1999. 'The Dark Side of Creativity', in *Encyclopedia of Creativity* (Vol. 1), ed. by Mark A. Runco and Steven R. Pritzker (San Diego, CA: Academic Press), pp. 483–91

McLaughlin, Brian, and Karen Bennett. 2021. 'Supervenience', in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta (https://plato.stanford.edu/archives/sum2021/entries/supervenience/; accessed 22/05/2022)

McMahon, Darrin M. 2013. *Divine Fury: A History of Genius* (New York: Basic Books)

Mednick, Sarnoff. 1962. 'The Associative Basis of the Creative Process', in *Psychological Review* 69(3), pp. 220–32

Mele, Alfred R. 1992. *Springs of Action: Understanding Intentional Behavior* (Oxford: Oxford University Press)

Mele, Alfred R. 2003. *Motivation and Agency* (Oxford: Oxford University Press)

Mill, John Stuart. 1998. ‘Utilitarianism’, in *Ethics: The Big Questions*, ed. by James P. Sterba (Oxford: Blackwell Publishers), pp. 119–33

Miller, Alexander, and Olivia Sultanescu. 2022. ‘Rule-Following and Intentionality’, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta (https://plato.stanford.edu/archives/sum2022/entries/rule-following/; accessed 15/06/2022)

Miller, Arthur I. 2019. *The Artist in the Machine: The World of AI-Powered Creativity* (Cambridge, MA: MIT Press)

Miller, Geoffrey. 2000. ‘Evolution of Human Music through Sexual Selection’, in *The Origins of Music*, ed. by Nils L. Wallin, Björn Merker, and Steven Brown (Cambridge, MA: MIT Press), pp. 329–60

Miller, Geoffrey. 2001. ‘Aesthetic Fitness: How Sexual Selection Shaped Artistic Virtuosity as a Fitness Indicator and Aesthetic Preferences as Mate Choice Criteria’, in *Bulletin of Psychology and the Arts* 2(1), pp. 20–25

Mitchell, Sandra D. 2000. ‘Dimensions of Scientific Law’, in *Philosophy of Science* 67(2), pp. 242–65

Mogensen, René. 2018. ‘Dynamic Concept Spaces in Computational Creativity for Music’, in *Philosophy and Theory of Artificial Intelligence 2017*, ed. by Vincent C. Müller (Cham: Springer), pp. 57–68

Moi, Toril. 2017. *Revolution of the Ordinary: Literary Studies after Wittgenstein, Austin, and Cavell* (Chicago: University of Chicago Press)

Monk, Ray. 1990. *Ludwig Wittgenstein: The Duty of Genius* (London: Vintage)

Moore, G. E. 1975. ‘Progress in Digital Integrated Electronics’, in *Technical Digest 1975: International Electron Devices Meeting*, pp. 11–13

More, Max. 2010. ‘The Overhuman in the Transhuman’, in *Journal of Evolution and Technology* 21(1), pp. 1–4

More, Max. 2013a. ‘The Philosophy of Transhumanism’, in *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, ed. by Max More and Natasha Vita-More (Chichester: Wiley-Blackwell), pp. 3–17

More, Max. 2013b. ‘A Letter to Mother Nature’, in *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, ed. by Max More and Natasha Vita-More (Chichester: Wiley-Blackwell), pp. 449f.

Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence* (Cambridge, MA; London: Harvard University Press)

Moravec, Hans. 1999. *Robot: Mere Machine to Transcendent Mind* (Oxford: Oxford University Press)

Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso)

Mulhall, Stephen. 2001. *Inheritance and Originality: Wittgenstein, Heidegger, Kierkegaard* (Oxford: Clarendon Press)

Mulhall, Stephen. 2014. *On Being in the World: Wittgenstein and Heidegger on Seeing Aspects* (London: Routledge)

Mumford, Michael D. 2003. 'Where have we been, where are we going? Taking Stock in Creativity Research', in *Creativity Research Journal* 15(2-3), pp. 107–20

Nabokov, Vladimir. 1981. *Lectures on Russian Literature*, ed. by Fredson Bowers (New York, NY: Harcourt / Bruccoli Clark Layman)

National Advisory Committee on Creative and Cultural Education. 1999. 'All Our Futures: Creativity, Culture and Education, Report to the Secretary of State for Education and Employment the Secretary of State for Culture, Media and Sport' (London: DfEE)

Netanel, Neil Weinstock. 2008. *Copyright's Paradox* (Oxford: Oxford University Press)

Neumaier, Otto. 1989. 'A Wittgensteinian View of Artificial Intelligence', in *Artificial Intelligence: The Case Against*, ed. by Rainer Born (London: Routledge), pp. 132–73

Nickles, Thomas. 2017. 'Scientific Revolutions', in The Stanford Encyclopedia of Philosophy, ed. by Edward N. Zalta (https://plato.stanford.edu/archives/win2017/entries/scientific-revolutions/; accessed 18/06/2019)

Nietzsche, Friedrich Wilhelm. 1996. *Human, All Too Human*, trans. by R. J. Hollingdale (Cambridge: Cambridge University Press)

Nietzsche, Friedrich. 1999. *The Birth of Tragedy and Other Writings*, trans. by Ronald Speirs, ed. by Raymond Geuss and Ronald Speirs (Cambridge: Cambridge University Press)

Nietzsche, Friedrich. 2005. 'Ecce Homo', in *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, trans. by Judith Norman, ed. by Aaron Ridley and Judith Norman (Cambridge: Cambridge University Press), pp. 69–152

Nietzsche, Friedrich. 2006. *Thus Spoke Zarathustra: A Book for All and None*, trans. by Adrian Del Caro, ed. by Adrian Del Caro and Robert B. Pippin (Cambridge: Cambridge University Press)

Nietzsche, Friedrich Wilhelm. 2007. *On the Genealogy of Morality*, trans. by Carol Diethe, ed. by Keith Ansell-Pearson (Cambridge; New York, NY: Cambridge University Press)

Novitz, David. 1999. 'Creativity and Constraint', in *Australasian Journal of Philosophy* 77(1), pp. 67–82

Novitz, David. 2003. 'Explanations of Creativity', in *The Creation of Art: New Essays in Philosophical Aesthetics*, ed. by Berys Gaut and Paisley Livingston (Cambridge: Cambridge University Press), 174–91

Nyíri, J. C. 1989. 'Wittgenstein and the Problem of Machine Consciousness', in *Grazer Philosophische Studien* 33, pp. 375–94

Ockham, William of. 1994. 'Five Questions on Universals from His *Ordinatio* d. 2, Qs. 4–8', in *Five Texts on the Mediaeval Problem of Universals: Porphyry, Boethius, Abelard, Duns Scotus, Ockham*, ed. and trans. by Paul Vincent Spade (Indianapolis, IN: Hackett), 114–231

O'Connor, Timothy. 1995. 'Agent Causation', in *Agents, Causes, and Events: Essays on Indeterminism and Free Will* (New York, NY; Oxford: Oxford University Press), pp. 173–200

O'Hear, Anthony. 1995. 'Art and Technology: An Old Tension', in *Royal Institute of Philosophy Supplement* 38, pp.143–58

Omohundro, Stephen M. 2008. 'The Basic AI Drives', in *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, ed. by Pei Wang, Ben Goertzel, and Stan Franklin (Amsterdam: IOS), pp. 483–92

Ortega y Gasset, José. 1946. *Mission of the University*, trans. by Howard Lee Nostrand (London: Kegan Paul)

Osborn, Thomas. 2003. 'Against "Creativity": A Philistine Rant', in *Economy and Society* 32(4), 507–25

Oxford English Dictionary. 1971. Compact Edition (Vol. 2) (Oxford: Oxford University Press)

Park, Michael, Erin Leahey, and Russell J. Funk. 2023. 'Papers and Patents are Becoming less Disruptive over Time', in *Nature* 613, pp. 138–44

Paul, Elliot Samuel, and Scott Barry Kaufman. 2014. 'Introducing: The Philosophy of Creativity', in *The Philosophy of Creativity: New Essays*, ed. by Elliot Samuel Paul and Scott Barry Kaufman (New York, NY: Oxford University Press), pp. 3–16

Paul, Elliot Samuel, and Dustin Stokes. 2018. 'Attributing Creativity', in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 193–209

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why* (London: Penguin Books)

Penco, Carlo. 2012. 'Updating the Turing Test: Wittgenstein, Turing and Symbol Manipulation', in *Open Journal of Philosophy* 2, pp. 189–94

Penrose, Roger. 1994. *Shadows of the Mind: A Search for the Missing Science of Consciousness* (Oxford: Oxford University Press)

Penrose, Roger. 1999. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford: Oxford University Press)

Pfeifer, Rolf, and Josh Bongard. 2007. *How the Body Shapes the Way We Think: A New View of Intelligence* (Cambridge, MA; London: MIT Press)

Plato. 1971. 'Timaeus', in *Timaeus and Critias*, trans. by Desmond Lee (Harmondsworth: Penguin), pp. 37–66

Plato. 1987. 'Ion', in *Early Socratic Dialogues*, ed. and trans. by Trevor J. Saunders (Harmondsworth: Penguin), pp. 37–66

Poincaré, Henri. 1910. 'Mathematical Creation', in *The Monist* 20(3), pp. 321–35

Pope, Alexander. 1876. *The Complete Poetical Works of Alexander Pope: With an Original Memoir of the Author Embracing Notices of Many Eminent Contemporaries, Critical and Explanatory Notes, Original and Selected, and Several Valuable Miscellaneous Productions, Not Contained in Any Other Edition* (New York, NY: World Publishing House)

Popper, Karl R. 1972. ‘Of Clouds and Clocks’, in *Objective Knowledge: An Evolutionary Approach* (Oxford: Clarendon Press), pp. 206–55

Popper, Karl R. 2002a. *Conjectures and Refutations: The Growth of Scientific Knowledge* (London; New York, NY: Routledge)

Popper, K. 2002b. *The Logic of Scientific Discovery* (London; New York, NY: Routledge)

Proudfoot, Devon, Aaron C. Kay, Christy Z. Koval. 2015. ‘A Gender Bias in the Attribution of Creativity: Archival and Experimental Evidence for the Perceived Association Between Masculinity and Creative Thinking’, in *Psychological Science* 26(11), pp 1751–61

Proudfoot, Diane. 2004a. ‘Robots and Rule-Following’, in *Alan Turing: Life and Legacy of a Great Thinker*, ed. by Christof Teuscher (Berlin: Springer), pp. 359–79

Proudfoot, Diane. 2004b. ‘The Implications of an Externalist Theory of Rule-Following Behaviour for Robot Cognition’, in *Minds and Machines* 14, pp. 283–308

Puttenham, George. 2016. *The Art of English Poesy: A Critical Edition*, ed. by Frank Whigham and Wayne A. Rebhorn (Ithaca, NY: Cornell University Press)

Quine, Willard Van Orman. 1951. ‘Two Dogmas of Empiricism’, in *The Philosophical Review* 60(1), pp. 20–43

Quine, Willard Van Orman. 1981. *Theories and Things* (Cambridge, MA; London: Belknap Press of Harvard University Press)

Quine, Willard Van Orman, and J. S. Ullian. 1978. *The Web of Belief* (New York, NY: Random House)

Ramsey, William M. 1997. ‘Do Connectionist Representations Earn Their Explanatory Keep?’, in *Mind & Language* 12(1), pp. 34–66

Ramsey, William M. 2007. *Representation Reconsidered* (Cambridge: Cambridge University Press)

Ramsey, William M., Stephen Stich, and Joseph Garon. 1990. ‘Connectionism, Eliminativism and The Future of Folk Psychology’, in *Philosophical Perspectives* 4, pp. 499–533

Raven, Michael. 2020. ‘Introduction’, in *The Routledge Handbook of Metaphysical Grounding*, ed. by Michael Raven (New York, NY: Routledge), pp. 1–14

Raz, Joseph. 2005. *The Practice of Value* (Oxford: Clarendon Press; New York, NY: Oxford University Press)

Reckwitz, Andreas. 2016. *Kreativität und Soziale Praxis: Studien zur Sozial- und Gesellschaftstheorie* (Bielefeld: Transcript Verlag)

Reckwitz, Andreas. 2017. *The Invention of Creativity: Modern Society and the Culture of the New* (Cambridge: Polity)

Reckwitz, Andreas. 2018. ‘The Creativity Dispositif and the Social Regimes of the New’, in *Innovation Society Today: Perspectives, Fields, and Cases*, ed. by Werner Rammert, Arnold Windeler, Hubert Knoblauch, and Michael Hutter (Wiesbaden: Springer Fachmedien Wiesbaden), pp. 127–45

Rees, Tobias. 2021. ‘Planetary Politics From Inside The Prison-House Of Language’, interviewed by Nils Gilman, in *Noema* (https://www.noemamag.com/planetary-politics-from-inside-the-prison-house-of-language/; accessed 16/06/2021)

Rees, Tobias. 2022. ‘Non-Human Words: On GPT-3 as a Philosophical Laboratory’, in *Daedalus* 151(2), pp. 168–82

Robertson, Ritchie. 2020. *The Enlightenment: The Pursuit of Happiness, 1680-1790* (London: Allen Lan)

Rogers, Carl R. 1970. ‘Towards a Theory of Creativity’, in *Creativity: Selected Readings*, ed. by Philip E. Vernon (Harmondsworth: Penguin Books), pp. 137–51

Roose, Kevin. 2022. ‘An A.I.-Generated Picture Won an Art Prize. Artists Aren’t Happy’, in *The New York Times*, 2 September 2022 (https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html; accessed 21/11/2022)

Rorty, Richard. 1965. ‘Mind-Body Identity, Privacy, and Categories’, in *Review of Metaphysics* 19(1), pp. 24–54

Rosenblum, Robert. 1996. ‘The Spanishness of Picasso’s Still Life’, in *Picasso and the Spanish Tradition*, ed. by Jonathan Brown (New Haven, CT: Yale University Press), pp. 61–93

Rosenkranz, Karl. 2015. *Aesthetics of Ugliness: A Critical Edition*, trans. by Andrei Pop and Mechtild Widrich (London: Bloomsbury)

Rousseau, Jean-Jacques. 1994. ‘The Social Contract’, in *Discourse on Political Economy and The Social Contract*, trans. by Christopher Betts (Oxford: Oxford University Press), pp. 43–168

Runco, Mark A., and, Albert, Robert S. 2010. ‘Creativity Research: A Historical View’, in *The Cambridge Handbook of Creativity*, ed. by James C. Kaufman and Robert J. Sternberg (Cambridge: Cambridge University Press), pp. 3–19

Russell, Bertrand. 1995. *The Analysis of Mind* (London: Routledge)

Russell, Stuart. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control* (New York, NY: Viking)

Rutherford, Adam. 2022. *Control: The Dark History and Troubling Present of Eugenics* (London: Weidenfeld & Nicolson)

Sacks, Oliver. 2007. *Musicophilia: Tales of Music and the Brain* (New York, NY; Toronto: Alfred A. Knopf)

Saint-Lambert, Jean-François de (ascribed). 2007. ‘Genius’, in *The Encyclopedia of Diderot & d’Alembert: Collaborative Translation Project*, trans. by John S.D. Glaus (Ann Arbor, MI: Michigan Publishing, University of Michigan Library) (https://quod.lib.umich.edu/d/did/did2222.0000.819; accessed 27/02/2022)

Sawyer, R. Keith. 2012. *Explaining Creativity: The Science of Human Innovation* (New York, NY: Oxford University Press)

Scaligero, Giulio Cesare. 1905. *Select Translations from Scaliger's Poetics*, trans. by Frederick Morgan Padelford (New York, NY: Henry Holt)

Schlegel, Friedrich. 1991. ‘Critical Fragments’, in *Philosophical Fragments*, trans. by Peter Firchow (Minneapolis, MN: University of Minnesota Press), pp. 1–16

Schlosser, Markus. 2019. ‘Agency’, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta (https://plato.stanford.edu/archives/win2019/entries/agency/; accessed 07/08/2020)

Schmidinger, Heinrich. 2008. ‘Das Auszeichnende des Menschen ist seine Kreativität: Eine geistesgeschichtliche Einstimmung’, in *Der Mensch – ein kreatives Wesen? Kunst – Technik – Innovation*, ed. by Heinrich Schmidinger and Clemens Sedmak (Darmstadt: Wissenschaftliche Buchgesellschaft), pp. 7–24

Schneider, Susan. 2019. *Artificial You: AI and the Future of Your Mind* (Princeton: Princeton University Press)

Schrift, Alan D. 2001. Nietz*sche’s Postmoralism: Essays on Nietzsche’s Prelude to Philosophy’s Future*, ed. by Richard Schacht (Cambridge: Cambridge University Press), pp. 47–62

Searle, John. 2007. *Freedom and Neurobiology* (New York, NY: Columbia University Press)

Seelig, Carl. 1986. *Helle Zeit – Dunkle Zeit: In Memoriam Albert Einstein* (Braunschweig; Wiesbaden: Vieweg)

Sejnowski, Terrence J. 2018. *The Deep Learning Revolution* (Cambridge, MA: The MIT Press)

Selinger, Evan. 2009. ‘Cyborgs’, in *A Companion to the Philosophy of Technology*, ed. by Jan Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendrick (Chichester, UK; Malden, MA: Wiley-Blackwell), pp. 154–56

Sepkoski Jr, J. John, Richard K. Bambach, David M. Raup, and James W. Valentine. 1981. ‘Phanerozoic Marine Diversity and the Fossil Record’, in *Nature* 293, pp. 435–37

Shaftesbury, Anthony Ashley Cooper, Earl of. 1999. *Characteristics of Men, Manners, Opinions, Times*, ed. by Lawrence Eliot Klein (Cambridge: Cambridge University Press)

Shanahan, Murray. 2015. *The Technological Singularity* (Cambridge, MA: The MIT Press)

Shanker, Stuart. 2002. *Wittgenstein’s Remarks on the Foundations of AI* (London; New York, NY: Routledge)

Shevlin, Henry. 2021. ‘Rethinking Creative Intelligence: Comparative Psychology and the Concept of Creativity’, in *European Journal for Philosophy of Science* 11(16), pp. 1–21

Sidney, Philip. 2002. *An Apology for Poetry*, ed. by Geoffrey Shepherd (Manchester: Manchester University Press)

Simonton, Dean Keith. 1999. *Origins of Genius: Darwinian Perspectives on Creativity* (New York, NY; Oxford: Oxford University Press)

Simonton, Dean Keith. 2009. ‘Creativity as a Darwinian Phenomenon: The Blind-Variation and Selective-Retention Model’, in *The Idea of Creativity*, ed. by Michael Krausz, Denis Dutton, and Karen Bardsley (Leiden; Boston, MA: Brill), pp. 63–8

Simonton, Dean Keith. 2013. ‘After Einstein: Scientific Genius is Extinct’, in *Nature* 493(7434), p. 602

Sloman, Aaron. 1984. ‘The Structure of the Space of Possible Minds’, in *The Mind and the Machine: Philosophical Aspects of Artificial Intelligence*, ed. by Steve Torrance (Chichester: Ellis Horwood), pp. 35–42

Sluga, Hans. 2011. *Wittgenstein* (Chichester: Wiley-Blackwell)

Sorensen, Roy. 2018. ‘Vagueness’, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta (https://plato.stanford.edu/archives/sum2018/entries/vagueness/; accessed 12/05/2020)

Sorgner, Stefan Lorenz. 2009. ‘Nietzsche, the Overhuman, and Transhumanism’, in *Journal of Evolution and Technology* 20(1), pp. 29–42

Sorgner, Stefan Lorenz. 2021. *We have Always been Cyborgs: Digital Data, Gene Technologies, and an Ethics of Transhumanism* (Bristol: Bristol University Press)

Sternberg, Robert J. 2003. *Wisdom, Intelligence, and Creativity Synthesized* (Cambridge: Cambridge University Press)

Stevenson, Leslie. 2003. ‘Twelve Conceptions of Imagination’, in *The British Journal of Aesthetics* 43(3), pp. 238–59

Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science* (Cambridge, MA: MIT Press)

Stokes, Dustin. 2008. ‘A Metaphysics of Creativity’, in *New Waves in Aesthetics*, ed. by Kathleen Stock and Katherine Thomson-Jones (Basingstoke: Palgrave Macmillan), pp. 105–24

Stokes, Dustin. 2011. ‘Minimally Creative Thought’, in *Metaphilosophy* 42(5), pp. 658–81

Stokes, Dustin. 2014. ‘The Role of Imagination in Creativity’, in *The Philosophy of Creativity: New Essays*, ed. by Elliot Samuel Paul and Scott Barry Kaufman (New York, NY: Oxford University Press), pp. 157–84

Strawson, Galen. 2004. ‘Free Agents’, in *Philosophical Topics* 32(1/2), pp. 371–402

Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics* (London: Methuen)

Stroud, Barry. 2000. ‘Mind, Meaning, and Practice’, in *Meaning, Understanding, and Practice* (Oxford: Oxford University Press), pp. 170–92

Stroud, Barry. 2012. ‘Meaning and Understanding’, in *Wittgenstein and the Philosophy of Mind*, ed. by Jonathan Ellis and Daniel Guevara (Oxford: Oxford University Press), pp. 19–36

Taliaferro, Charles, and Meredith Varie. 2018. ‘The Active and Passive Life of Creativity: An Essay in a Platonic Key’, in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London; New York, NY: Routledge), pp. 140–51

Taylor, Charles. 1991. *The Ethics of Authenticity* (Cambridge, MA; London: Harvard University Press)

Taylor, Charles. 1995. ‘To Follow a Rule’, in *Philosophical Arguments* (Cambridge, MA; London: Harvard University Press), pp. 165–80

Tegmark, Max. 2018. *Life 3.0: Being Human in the Age of Artificial Intelligence* (London: Penguin Books)

Thagard, Paul, and Terrence C. Stewart. 2011. ‘The AHA! Experience: Creativity Through Emergent Binding in Neural Networks’, in *Cognitive Science* 35(1), pp. 1–33

Thagard, Paul. 2012. ‘Creative Combination of Representations: Scientific Discovery and Technological Invention’, in *Psychology of Science: Implicit and Explicit Processes*, ed. by Robert W. Proctor and E.J. Capaldi (Oxford: Oxford University Press), pp. 389–405

’t Hooft, Gerard. 2016. ‘The Evolution of Quantum Field Theory: From QED to Grand Unification’, in *The Standard Theory of Particle Physics: Essays to Celebrate CERN’s 60th Anniversary*, ed. by Luciano Maiani and Luigi Rolandi (Singapore: World Scientific Publishing Company), pp. 1–27

Turing, Alan M. 1937. ‘On Computable Numbers, with an Application to the Entscheidungs-problem’, in *Proceedings of the London Mathematical Society* 42(1), pp. 230–65

Turing, Alan M. 1950. ‘Computing Machinery and Intelligence’, in *Mind* 59(236), pp. 433–60

Turing, Alan M. 1986. ‘Lecture to the London Mathematical Society on 20 February 1947’, in *A. M. Turing’s Ace Report of 1946 and Other Papers*, ed. by Brian E. Carpenter and Robert W. Doran (Cambridge, MA: MIT Press), pp. 106–24

Turner, Mark. 2010. ‘Conceptual Integration’, in *The Oxford Handbook of Cognitive Linguistics*, ed. by Dirk Geeraerts and Herbert Cuyckens (New York, NY: Oxford University Press), pp. 377–93

Turner, Mark, and Gilles Fauconnier. 1999. ‘A Mechanism of Creativity’, in *Poetics Today* 20(3), pp. 397–418

Tyson, Neil deGrasse. 2007. *Death by Black Hole: And Other Cosmic Quandaries* (New York, NY; London: W. W. Norton)

Tzafestas, Spyros G. 2022. *An Introduction to Robophilosophy: Cognition, Intelligence, Autonomy, Consciousness, Conscience, and Ethics* (Gistrup: River Publishers)

Ulam, Stanislaw. 1958. ‘Tribute to John von Neumann’, in *Bulletin of the American Mathematical Society* 64(3, part 2), pp. 1–49

Van Gulick, Robert. 2022. ‘Consciousness’, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta & Uri Nodelman (https://plato.stanford.edu/archives/win2022/entries/consciousness/; accessed 20/12/2022)

Vasari, Giorgio. 1987. *The Lives of the Artists* (Vol. 1), trans. by George Bull (Harmondsworth: Penguin)

Vattimo, Gianni. 2014. ‘The Structure of Artistic Revolutions’, in *Postmodernism: A Reader*, ed. by Thomas Docherty (London; New York, NY: Routledge), pp. 110–19

Vermeulen, Andreas François. 2020. *Industrial Machine Learning: Using Artificial Intelligence as a Transformational Disruptor* (New York, NY: Apress)

Vinge, Vernor. 1993. 'The Coming Technological Singularity: How to Survive in the Post-Human Era', in *NASA Lewis Research Center, Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, pp. 11–22 (https://ntrs.nasa.gov/citations/19940022856; accessed 03/04/2022)

Vygotsky, Lev Semenovich. 2004. 'Imagination and Creativity in Childhood', in *Journal of Russian and East European Psychology* 42(1), pp. 7–97

Wallace, R. Jay. 2005. 'Introduction', in Joseph Raz: *The Practice of Value* (Oxford: Clarendon Press; New York, NY: Oxford University Press), pp. 1–12

Wallas, Graham. 1926. *The Art of Thought* (New York, NY: Harcourt, Brace & Company)

Walton, Kendall L. 1990. *Mimesis as Make-Believe* (Cambridge, MA: Harvard University Press)

Ward, Adolfus William. 1875. *A History of English Dramatic Literature: To the Death of Queen Anne* (Vol. 1) (London: Macmillan and Co.)

Warren, Jared, 2020, 'Killing Kripkenstein's Monster', in *Noûs* 54(2), pp. 257–89

Watson, James D., and Francis H. C. Crick. 1953. 'Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid', in *Nature* 171, pp. 737–38

Weibel, Peter. 2019. 'Wider das moderne Paradigma romantischer Entgrenzung und für eine neue Ethik im digitalen Zeitalter', in *Grenzen und Entgrenzung: Ethische Orientierung in einer destabilisierten Welt*, ed. by Hermes Andreas Kick and Manfred Oeming (Berlin: Lit Verlag), pp. 123–32

Weiner, Robert Paul. 2000. *Creativity and Beyond: Cultures, Values, and Change* (Albany, NY: State University of New York Press)

Wheeler, Michael. 2008. 'Cognition in Context: Phenomenology, Situated Robotics and the Frame Problem', in *International Journal of Philosophical Studies* 16(3), pp. 323–49

Wheeler, Michael. 2018. 'Talking about more than Heads: The Embodied, Embedded and Extended Creative Mind', in *Creativity and Philosophy*, ed. by Berys Gaut and Matthew Kieran (London: Routledge), pp. 230–50

Wilczek, Frank. 2020. 'The Unity of Intelligence', in *Possible Minds: Twenty-Five Ways of Looking at AI*, ed. by John Brockman (Penguin Books: Penguin Books), pp. 64–75

Williams, Bernard. 1973. 'The Makropolus Case: Reflections on the Tedium of Immortality', in *Problems of the Self: Philosophical Papers 1956–1972* (Cambridge: Cambridge University Press), pp. 82–100

Williams, F.C. 1975. 'Early Computers at Manchester University', in *The Radio and Electrical Engineer* 45, pp. 327–31

Williams, Raymond. 1961. *The Long Revolution* (London: Penguin Books)

Williamson, Timothy. 1994. *Vagueness* (London: Routledge)

Williamson, Timothy. 1997. ‘Précis of Vagueness’, in *Philosophy and Phenomenological Research* 57(4), pp. 921–28

Williamson, Timothy. 2013. ‘Wittgenstein’s Metaphilosophy, by Paul Horwich’, in *European Journal of Philosophy* 21(S2), pp. e7–e10

Winckelmann, Johann Joachim. 2013. ‘Thoughts on the Imitation of Greek Works in Painting and the Art of Sculpture’, in *Johann Joachim Winckelmann on Art, Architecture, and Archaeology*, trans. by David R. Carter (Rochester, NY: Camden House), pp. 31–56

Winner, Langdon. 1986. *The Whale and the Reactor: A search for Limits in an Age of High Technology* (Chicago, IL: University of Chicago Press)

Wittgenstein, Ludwig. 1964. *Blue Book* (Oxford: Basil Blackwell)

Wittgenstein, Ludwig. 1966. *Lectures & Conversations on Aesthetics, Psychology, and Religious Belief*, ed. by Cyril Barrett (Oxford: Basil Blackwell)

Wittgenstein, Ludwig. 1967. *Zettel*, ed. by Gertrude Elizabeth Margaret Anscombe and Georg Henrik von Wright, trans. by Gertrude Elizabeth Margaret Anscombe (Oxford: Blackwell)

Wittgenstein, Ludwig. 1975. *Philosophische Untersuchungen* (Frankfurt am Main: Suhrkamp)

Wittgenstein, Ludwig. 1978. *Remarks on the Foundations of Mathematics*, ed. by Gertrude Elizabeth Margaret Anscombe, Rush Rhees, and Georg Henrik von Wright, trans. by Gertrude Elizabeth Margaret Anscombe (Oxford: Blackwell)

Wittgenstein, Ludwig. 1980a. *Culture and Value*, trans. by Peter Winch (Chicago, IL: University of Chicago Press)

Wittgenstein, Ludwig. 1980b. *Remarks on the Philosophy of Psychology* (Vol. 1), ed. by Gertrude Elizabeth Margaret Anscombe and Georg Henrik von Wright, trans. by Gertrude Elizabeth Margaret Anscombe (Oxford: Blackwell)

Wittgenstein, Ludwig. 1986. *Philosophical Investigations*, trans. by Gertrude Elizabeth Margaret Anscombe (Oxford: Basil Blackwell)

Wittgenstein, Ludwig. 2016a. *Lectures, Cambridge, 1930-1933: From the Notes of G. E. Moore*, ed. by David G. Stern, Brian W. Rogers, and Gabriel Citron (Cambridge: Cambridge University Press)

Wittgenstein, Ludwig. 2016b. *On Certainty* (Oxford: Blackwell Publishing)

Wright, Crispin. 1980. *Wittgenstein on the Foundations of Mathematics* (London: Duckworth)

Wundt, Wilhelm Max. 1874. *Grundzüge de Physiologischen Psychologie* (Vol. 1) (Leipzig: W. Engelmann)

Ye, Andre. 2020. ‘Real Artificial Intelligence: Understanding Extrapolation vs Generalization’ (https://towardsdatascience.com/real-artificial-intelligence-understanding-extrapolation-vs-generalization-b8e8dcf5fd4b; accessed 06/07/2021)

Yudkowsky, Eliezer. 2008. ‘Artificial Intelligence as a Positive and Negative Factor in Global Risk’, in *Global Catastrophic Risks*, ed. by Nick Bostrom and Milan M. Ćirković (New York, NY: Oxford University Press), pp. 308–45

Zagzebski, Linda Trinkaus. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press)

Zola, Émile. 1894. 'The Experimental Novel', in *The Experimental Novel and Other Essays*, trans. by Belle M. Sherman (New York (State): The Cassell Publishing Co), pp. 1–56

Zweig, Stefan. 2006. *Chess Story* (New York, NY: New York Review Book Classics)