PhD-FHSE-2021-025
The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 12/07/2021 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN *PHILOSOPHIE*

by

Hannes FRAISSLER
Born on 29 January 1986 in Voitsberg (Austria)

THOUGHT, LANGUAGE, AND REASONING:
PERSPECTIVES ON THE RELATION BETWEEN MIND AND LANGUAGE

Dissertation defence committee

Dr Frank Hofmann, dissertation supervisor
Professor, Université du Luxembourg

Dr Mark Textor
Professor, King's College London

Dr Dietmar Heidemann, Chairman
Professor, Université du Luxembourg

Dr Holger Sturm
Professor, Universität des Saarlandes

Dr Thomas Raleigh, Vice Chairman
Associate Professor, Université du Luxembourg

Abstract

This dissertation is an investigation into the relation between mind and language from different perspectives, split up into three interrelated but still, for the most part, self-standing parts. Parts I and II are concerned with the question how thought is affected by language while Part III investigates the scope covered by mind and language respectively.

Part I provides a reconstruction of Ludwig Wittgenstein's famous Private Language Argument in order to apply the rationale behind this line of argument to the relation between mind and language. This argumentative strategy yields the conclusion that reasoning – an important type of thought – is constitutively dependent on language possession and is therefore not available to non-linguistic creatures. This result is achieved by considering the preconditions for reasoning – given that it is a rule-governed activity – and eliminating competitors to language for providing reasoners with what it takes to reason.

Part II provides a critical outlook on the wide and highly heterogeneous field of linguistic relativity theories. It is argued that no kind of linguistic relativity whatsoever follows from the conclusion of Part I – i.e., the claim that reasoning is constitutively dependent on having a language. While Part II does not provide a conclusive argument against the linguistic relativity hypothesis, it is argued that endorsement of linguistic relativity theories often rests on a mistaken assumption to the effect that language and culture are interwoven in a way which makes it impossible to separate culture and language, as well as their respective studies. This assumption is undermined by providing examples of languages which clearly predate their culture (Esperanto) or do not even have a culture at all (Klingon). So, the assumption that language and culture are inextricably intertwined is refuted by way of counterexample.

Part III provides an in-depth examination of the Principle of Expressibility – prominently endorsed and formulated by John Searle – which claims that whatever can be thought can also be said. The domains of what can be thought and of what can be said are considered in set theoretic terms in order to determine whether one is contained in the other, so that everything we can think can also be adequately communicated. After thorough study of interpretative issues regarding the Principle of Expressibility and consideration of the most pressing potential counterexamples to the principle, we can conclude that we have good reason to believe in the truth of the Principle of Expressibility. In conclusion, the achieved results are related back to prominent positions in the discussion about thought and language which already make their appearance in the very beginning of this investigation. The final chapter of this dissertation reminds us that eminent figures in philosophy have often taken a wrongheaded perspective on the relation between language and thought, so that language has frequently appeared to be an impediment to thought. We can, however, confidently conclude that language, on the contrary, is by far our most apt means for thought and that reasoning would not even be possible without the resources language provides.

Thought, Language, and Reasoning

Perspectives on the Relation Between Mind and Language

Hannes Fraissler
University of Luxembourg

For my family

To those who have always been with me,
as well as those who cannot be with us anymore

This document was composed
using freely available
open source software:

LyX (text editing)
`https://www.lyx.org/`

MiKTeX (LaTeX distribution)
`https://miktex.org/`

Zotero (reference management)
`https://www.zotero.org/`

SumatraPDF (document viewer)
`https://www.sumatrapdfreader.org/`

Inkscape (image vectorization)
`https://inkscape.org/`

Preface and Acknowledgments

This is an investigation into the relation between thought and language and, as indicated in the thesis's subtitle, it covers more than one perspective on the topic. The different perspectives provided roughly align with the three parts of the investigation at hand and represent topics and approaches which have kept my mind busy for many years past. This situation will probably not change in the near future since I have opened some doors in the course of this investigation which could probably occupy a philosopher for a lifetime. So, I certainly do not wish to claim that there is nothing more to say about the topics treated here (even from my part). I do hope, on the contrary, for many occasions to further investigate considerations and arguments I merely sketched in what is to follow, as well as for occasions to elaborate on one or the other topic that might give rise to new debate – if things go well.

The general outlook of this thesis is in good agreement with the recently evolving research program of Cultural Evolutionary Psychology (cf. Heyes, 2018), which is open to the view that reasoning is a cognitive gadget – dependent "[. . .] on cognitive mechanisms constructed in the course of development through social interaction [. . .]" (Heyes, 2018, p. 220) – which in turn depends on another cognitive gadget: language. However, the claims made on the following pages are, as far as I can see, also compatible with the position that reasoning and language might be cognitive instincts rather than cognitive gadgets in Heyes's sense. Another congenial contemporary approach I think to have found in Mercier & Sperber's impressive study of *The Enigma of Reason*, which locates the "normal" environment for a proper functioning of reason "[. . .] in the midst of a discussion, as people exchange arguments and justifications with each other" (Mercier & Sperber, 2017, p. 10). I may add

that it is not a mere (evolutionary) coincidence that such discussions take place in language, for language – as I argue – is the very medium which makes reasoning possible in the first place.

So much for points of contact with recent developments, but there is also at least one notable breaking point with current and traditional approaches to the topic at hand. Some readers will certainly be surprised by the fact that I have virtually nothing to say about concept acquisition in this investigation and very little about concepts in general. This, I gather, will strike many as quite unusual – at least for an investigation of such extent – since especially concept acquisition arguably constitutes a standard topic to be covered by studies of the relation between thought and language. I presently follow a quite different approach to show that language enables reasoning and is therefore a crucial cognitive factor. Not all mental abilities are, of course, dependent on language. But reasoning – which is undoubtedly an absolutely central (although not defining)[1] human capacity – could not be available without language due to conceptual requirements which make reasoning different from mere thinking.

In any case, I hope that a missing treatment of concept acquisition will not be perceived as a lacuna of the present investigation. To the contrary, not only what I discuss here but also that it is possible to arrive at substantial claims without covering concept acquisition will hopefully prove to be a fruitful contribution to the discussion about mind, language, and their intricate interrelation. I will, for the most part, stay neutral regarding competing theories of mental representation and format of reasoning and thinking, as well as regarding the correct semantic theory for language. Although I have my preferences in this regard and although I think that some theories align more naturally with the claims defended here than others, I have tried to presuppose as little as possible in terms of the correct theory of the mind and the correct theory of language.

[1] I take the term "human" to be a name for the biological species we all belong to. This means that what makes a human being is simply the possession of a certain kind of DNA, granted a rough-and-ready version of biological essentialism. This presumably excludes all kinds of cognitive or cultural abilities from a definition of what makes an individual human.

Some remarks on the origins of the following investigation's individual parts are needed at this point. The core aspects of the material covered in Part I, and a few fractions of Part II, were presented at the *2nd Meeting on Cognition & Language – 2eC&L*, held at the Federal University of Uberlândia – UFU (Uberlândia, MG, Brazil), and during the international online research workshop *Mind and Language,* hosted by the University of Luxembourg. I wish to thank all discussants at both events for their valuable input which informed and greatly improved my (2021), where the aforementioned pieces of the current investigation have already appeared in print.

The core of the considerations presented in Part II were originally written not only as a part of the present investigation but also in order to be presented at a *PhD Conference in Social Sciences* at the University of Luxembourg. The talk, alas, never happened because unfortunate circumstances kept me from taking advantage of the opportunity to present the pertinent considerations to an audience which certainly would have provided quite valuable insights since I suspect that at least some highly esteemed colleagues from the social sciences and neighboring disciplines who participated in this event might have been sympathetic to the kind of position I attack. It is a pity, and I truly regret, that I missed the chance to "test" the lines of argument presented in Part II in front of a convenient target audience. I hope that I can make up for this lost opportunity in the future.

The content of Part III is loosely based on considerations which already appear in a (rather aged) term paper of mine. The paper was never published but written and handed in for a seminar about philosophy of language, held at the *Institute of Philosophy of the Catholic Faculty* at the University of Graz (Austria). Part III constitutes a considerably enlarged and revised version of the aforementioned paper.

Among the members of my family – biological as well as adopted – to whom this thesis is dedicated, I especially wish to express gratitude to my sister, who was there to take care of me while a major part (perhaps even the majority) of this thesis was initially written, despite the fact that she

lives a good thousand kilometers away. The friends and colleagues I owe acknowledgments are countless, or at least too numerous to mention them all.[2] I will therefore make only one exception for my dear colleague Deven Burks who not only discussed several topics of interest for this investigation with me but also proofread the entire manuscript in incredible detail and at a breathtaking pace. At least the linguistic quality of this thesis would not be what it is without his invaluable support. Last but definitely not least, I wish to thank my supervisor Frank Hofmann who kept me going even when – and sometimes especially when – I could not see light at the end of the tunnel anymore. Without his encouragement, this investigation would probably have petered out without ever seeing the light of the day. I therefore probably owe him, besides other things, the rest of my academic career. This will not be too small a debt, I hope.

[2] I can only assure every single person who deserves to be mentioned here that I had you all in mind when I wrote these lines.

Contents

Part I

Can We Think Without Language?

A Private Language Argument to Elucidate the Relation Between Language and Mind

Chapter 1

Private Language, Thought, and Reasoning

1.1 Introduction

When we ask about the relation between thought and language, or language and the mind, we can demarcate possible answers to this question along a spectrum. At one end of this spectrum we find a position which has come to be known as the *conduit metaphor* (Reddy, 1993) or the *communicative conception* of language (Carruthers & Boucher, 1998, p. 1). Although there are prominent voices to the contrary – e.g., Carruthers (2002, p. 657) and Kramsch (1998, p. 21) – I think it is fair to characterize this position as the "common sense" view about how thought and language are related (cf. Ahearn, 2017, p. 8; and Carruthers & Boucher, 1998, p. 1). It roughly tells us that thought is independent of language since language merely serves to communicate our language-independently premolded thoughts to others. In order to communicate, we need to "translate"[1] our thoughts into a public

[1] I put "translate" in scare quotes here because we usually only translate from one language into another, not from a non-linguistic into a linguistic medium. I also presuppose that, according to this view, there is no Mentalese or language of thought since thought would not be independent of language otherwise. If we think in Mentalese, i.e., the language of thought, the claim that thought is constitutively dependent on language becomes trivial. I will therefore not assume that there is a language of thought. However, even if there should be a language of thought, I will argue that a *public* language is necessary for

language. Yet, language has no impact whatsoever on (the structure of) thought because thinking is primary to and independent of language.[2]

At the other end of the spectrum we find various positions which fall under the umbrella term *"linguistic relativity."* Proponents of linguistic relativity theories often appeal to the early 20th century linguists Edward Sapir and, especially, Benjamin Lee Whorf and base their accounts on what has come to be known as the *Sapir-Whorf hypothesis*.[3] Different versions of the theory postulate different kinds and intensities of linguistic impact on thinking. Still, all varieties of linguistic relativity claim a substantial effect of language on thought. According to radical versions of the theory, the language we speak essentially determines how we perceive reality, and the thoughts of speakers of structurally extremely different languages become basically incommensurable.[4]

I am convinced that truth is not to be found at either extreme of this spectrum, but it is one thing to claim that we need to find a tenable middle ground and another thing entirely to carve out and argue for a convincing position. My aim in this first part of the present investigation is to demonstrate that language does indeed have a substantial impact on the mind – since a certain kind of thought, namely reasoning, constitutively depends on language – but no form of linguistic relativity whatsoever follows from this.

The line of argument I wish to present is closely related to Ludwig Wittgenstein's famous *Private Language Argument*. I will start (in section 1.2.1) with a short reminder of what Wittgenstein's Private Language Argument is before I provide (in section 1.2.2) a reconstruction of Wittgen-

reasoning nevertheless.

[2] See also Kramsch (2004, p. 250), who calls this position the "Cartesian view" and claims that it was implicitly adopted by both Western psychology and linguistics. Also Carruthers & Boucher (1998, p. 1) agree that "[t]he communicative conception [of language] is now dominant in many areas of the cognitive sciences (understood broadly to include cognitive psychology, empirical-minded philosophy of mind, linguistics, artificial intelligence – AI – and cognitive neuroscience) [...]" which plausibly makes this position not only the common sense view but also the academic mainstream view.

[3] The literature about linguistic relativity and the Sapir-Whorf hypothesis is extremely vast, but for a good and rather recent overview see Reines & Prinz (2009).

[4] For more on linguistic relativity, see Part II.

stein's argument, which I call *"PLA"*.[5] In a next step, I will (in section 1.3) sketch a distinction between thinking and reasoning, which needs to be drawn among conscious cognitive processes, which I wholesale refer to as "thought." This distinction is necessary before Wittgenstein's rationale – or at least what I take it to be – can be applied to the domain of mind and language. This is what I will do in section 2.1, where I present a *Private Reasoning Argument – PRA,* for short – which is constructed in close analogy with *PLA,* i.e., my reconstruction of Wittgenstein's Private Language Argument. If *PRA*, the Private Reasoning Argument, is cogent, it gives a direct – although not exhaustive – answer to the question of how mind and language are related. *PRA* rests on a controversial premise, which will be defended in section 2.2. Section 2.3 will be concerned with carving out the notion of language more clearly. I will then address a prominent concern about Wittgenstein's argument – namely verificationism – in section 3.1. Since I heavily rely on and basically just extend Wittgenstein's Private Language Argument to cover the relation between mind and language, my view may be accused of being verificationist as well, just as Wittgenstein's was. I will deal with this accusation, as far as it is directed against endorsement of *PRA*, in section 3.1; I will address verificationism more generally only much later in section 8.3 of Part III. Finally, as far as Part I is concerned, section 3.2 spells out some consequences of *PRA,* and section 3.3 investigates how reasoning and reasoner relate to each other before the big picture of the considerations presented in Part I is quickly summarized in section 3.4.

1.2 The *Private Language Argument*

While Ludwig Wittgenstein's *Private Language Argument* tells us that a private language cannot exist, it also points us towards the conclusion that

[5] My reconstruction of Wittgenstein's argument will be referred to as *"PLA"* throughout this text, whereas my usage of the expression "Private Language Argument" will indicate that I am not talking about my particular interpretation and reconstruction of the argument but rather about the argument more generally, usually disregarding exegetical differences in the interpretation of what has come to be known as Wittgenstein's Private Language Argument.

private reasoning cannot exist because reasoning is dependent on language. I am not the first to recognize this connection between considerations regarding private language and the relation between language and the mind. Donald Davidson (1991, p. 157) had already suggested an extension of Wittgenstein's rationale to the domain of mind and language. Yet I find Davidson's account rudimentary and defective in several ways, which will be addressed in section 3.1.[6]

However, to understand why Wittgenstein's considerations point us towards the conclusion that private reasoning cannot exist, we must first unpack the Private Language Argument. Only then can the aforementioned consequence be drawn and contextualized in the debate about mind and language. For the purpose at hand, there is no need for deep exegetical investigation into §§ 243-315 of Wittgenstein's (2009) *Philosophical Investigations*, which are commonly held to comprise the Private Language Argument. Any attempt to provide such would probably exceed the limits of this investigation by far, given the incredibly vast literature which has been sparked by Wittgenstein's Private Language Argument. Thomas Raleigh (2019, p. 70) even thinks that "[...] Wittgenstein's 'private language argument' probably has a fair claim to be the single most discussed passage of philosophy written in the twentieth century. [Footnote omitted]"

On the basis of this plausible assessment, it is no great surprise that there is even disagreement regarding the precise location of the Private Language Argument. Gordon Baker (1998, p. 325), for example, claims that the Private Language Argument contains "at least the remarks from § 243 to *§ 326*" (emphasis added). So he counts eleven more paragraphs as belonging to the Private Language Argument than is usually done. The following short intro-

[6] As a quick outlook on what is to come, we can state already that Davidson (2001) sketches the argument as applying to the entire realm of thought. I, in contrast, take it to be implausibly restrictive to limit the domain of creatures which are capable of having thoughts and beliefs to only those who also have language. Davidson's notion of thought is just too demanding to accommodate intuitively plausible belief ascriptions to non-linguistic animals. We therefore need to distinguish reasoning from thinking (which will be done in section 1.3) in order to successfully apply Davidson's suggestion to the former without excessively limiting the ascribability of the latter. This, as already mentioned, will be discussed in more detail in section 3.1.

duction to the Private Language Argument is just meant to provide a rough – some might even say naïve – outline of the argument in order to get a clear enough grasp on its functioning to follow the subsequent reconstruction (*PLA*) in section 1.2.2 and the later application to the relation between mind and language (*PRA*) in section 2.1.

So, issues about, e.g., Saul Kripke's (1982) interpretation of Wittgenstein's reflections on rule-following as a skeptical challenge (sometimes going under the portmanteau "Kripkenstein") will be left aside, just like questions about how faithful certain reconstructions of the Private Language Argument (especially *PLA*) stay to Wittgenstein's original intent. None of these exegetical issues will be of great importance here, and I am willing to accept any accusation of not having represented Wittgenstein's argument accurately. The only important question at hand is whether the following sketch and the subsequent reconstruction (*PLA*), inspired by what came to be called "the Private Language Argument," provide a cogent line of reasoning, which shall then be brought to service for further argumentative aims later on in this investigation – especially in chapter 2.

In order to roughly position my reading of the relevant paragraphs in Wittgenstein (2009), I would nonetheless be willing to state that I loosely follow Norman Malcolm's (1954) take on Wittgenstein's reasoning, which I even dare to call the standard or traditional interpretation of the Private Language Argument. But be that as it may. Since the reconstruction I present below (i.e., *PLA*) is meant to be a deductively valid argument with the conclusion that a private language is impossible, my take on Wittgenstein's Private Language Argument certainly qualifies as a representative of what is usually called "[. . .] the orthodox approach to private language" (Stern, 2011, p. 334). I will sketch the basic rationale of the Private Language Argument in the following section 1.2.1, and subsequently attach a reconstruction thereof in explicitly canonical form[7] in section 1.2.2.

[7] For what I mean by "canonical form," see p. 10.

1.2.1 General Depiction of the Private Language Argument

A language is private, in Wittgenstein's sense, if it principally cannot be learned or understood by anyone but a single speaker. This is (at least partly) because the expressions of a private language would refer to a speaker's private experiences, which are essentially inaccessible to anyone but the speaker. This is, at least, what the beginning of Wittgenstein's discussion of the private language (cf. Wittgenstein, 2009, § 243) and his famous example of the diary case (cf. Wittgenstein, 2009, § 258) suggest.[8]

Note that, in the subsequent reconstruction of the Private Language Argument in section 1.2.2, no use whatsoever is made of the alleged privacy of what is designated by a linguistic expression. This is an advantage because *PLA* is therefore open for and applicable to other conceivable forms of language, which are also private, albeit for different reasons than speaking about private sensations (cf. Bertolet, 1999, p. 741). It might be possible to read Wittgenstein (2009, §§ 259 and 269) as encouraging such a wider notion of private language since these passages suggest that it is not so much the privacy of what expressions of the private language mean that is decisive. Rather, the privacy of rules and criteria for correct use of such a language would make it private, regardless of what its expressions mean. But let us stick with the privacy of the designated experiences for the moment: Since no one but the speaker could possibly know the reference/meaning of these expressions, no one but the speaker could consequently understand the language.

Language is here conceived of as (*inter alia*) a system of rules, and therefore mastering a language is mastering its rules. This presupposes intersubjectivity because, without publicly available standards for determining when a rule is followed or not, the entire institution of rule-following breaks down. In other words, without publicly available standards, everything that seems right for a speaker will be right. Without a distinction between what merely

[8] Further passages which suggest that the privacy Wittgenstein is primarily concerned with – i.e., the reason why a private language is private – has to do with the privacy of experiences are Wittgenstein (2009, §§ 256 f, 268, 275, and 293).

seems right and what actually is right, any distinction between right and not right gets abandoned as well.

A private language would not allow for the distinction between actually following a rule and the mere impression of following a rule because there is in principle no way for a speaker to double-check whether she did indeed follow a rule correctly or whether it just seems to her that she did. But if this distinction between actually and allegedly following a rule is not available for a speaker, we are not even dealing with a language at all. Mastering a language comes down to mastering a set of rules. Yet, without any possibility to discriminate between successful and unsuccessful applications of a rule, there is no mastering of rules. Therefore, every possible language must be a public language.

To illustrate this idea, Wittgenstein (2009, § 258) gives us the example of someone who writes down a certain mark – let us say the sign "S" – in her diary whenever she has a certain sensation. The diarist wishes to adopt the following rule: "Whenever I have a sensation of type x, I will write down the sign 'S' in my diary." It is a crucial assumption in Wittgenstein's example that "S" does not belong to a public language since it is meant to refer to the private sensation of the diarist. Given that there are no publicly available correctness conditions for when the diarist is supposed to write down "S" – because only she can know whether she experiences an instance of x – it is not possible to make a proper distinction between cases where the diarist correctly followed the rule and when it merely seemed to the diarist that she correctly followed the rule.

Whatever seems right to the diarist will be right. Without any possibility to check for the correctness of her application of the rule, the diarist would need to follow the rule in private. But insofar as privately following a rule does not permit a proper distinction between correct and incorrect applications of a rule, we cannot follow a rule privately. There is nothing like private rule-following since it must always be possible – at least in principle – to make a distinction between correct and incorrect applications of a rule. Without this fundamental possibility there is no rule-following and *eo ipso* no language.

1.2.2 Reconstruction of the Private Language Argument – *PLA*

Since explicit reconstructions of the Private Language Argument seem to be surprisingly rare in the literature, what I provide here is my own reconstruction which explicitly marks assumptions and inferential relations among individual propositions (i.e., what I call "canonical form"). The only other reconstruction of the Private Language Argument which also provides these features is, to the best of my knowledge, Wrisley (2011, pp. 353 f). But his reconstruction sticks, unfortunately, too closely to the original text and is therefore too convoluted to carve out the underlying rationale in a helpful way. In an attempt to explicitly state the Private Language Argument in canonical form, I propose the following reconstruction:

***PLA*:** The *Private Language Argument* reconstructed [9]

$L1$	*	A (natural) language is (*inter alia*) a system of rules that can be mastered.
$L2$	*	A (natural) language can only be mastered if its rules can be mastered.
$L3$	*	Rules can only be mastered if it is possible to draw a distinction between correct and incorrect applications of a rule.
$L4$		A (natural) language can only be mastered if it is possible to draw a distinction between correct and incorrect applications of its rules. [via hypothetical syllogism from $L2$ and $L3$]
$L5$	*	A private language would not allow drawing the distinction between correctly and incorrectly applied rules.
$L6$		A private language cannot be mastered. [via modus tollens from $L4$ and $L5$]
$L7$	$\therefore$	A private language is not a possible (natural) language. [via modus tollens from $L1$ and $L6$]

[9] The asterisk (*) marks assumptions.

$L8$ $\therefore$ Every possible language is a public (i.e., non-private) language. [via contraposition from $L7$]

Premises $L1$ and $L2$ should be rather unproblematic. What they claim can count as "[...] a familiar view in philosophy and linguistics [...]" (Searle, 2011, p. 12). So, this way of thinking about language is well established and should not be overly controversial. The gist of Wittgenstein's reasoning is contained in premises $L3$ and $L5$, which are the critical assumptions needed to get *PLA* working. The rationale behind those crucial premises is the claim that we cannot follow a rule in private, i.e., the *rule-following constraint*, as explained in section 1.2.1. Since this principle will be referred to repeatedly on the following pages, we should state it again in a sufficiently prominent position.

The rule-following constraint: A rule cannot be followed in private.[10]

1.3 Thinking vs. Reasoning

A variation of this argument yields the conclusion that private reasoning is just as impossible as a private language, granted, of course, that *PLA* is successful. As a first step to apply *PLA*'s rationale to the relation between language and the mind, we need to make a distinction between two proper subsets of *thought* in a rather broad understanding.[11] The distinction is between *reasoning* on the one hand and all kinds of *thinking* which are not

[10] Wittgenstein (2009, § 202) states the matter as follows:

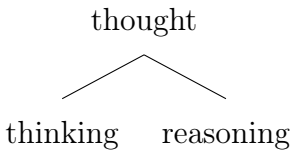
> That's why 'following a rule' is a practice. And to *think* one is following a rule is not to follow a rule. And that's why it's not possible to follow a rule 'privately'; otherwise, thinking one was following a rule would be the same thing as following it.

The corresponding passage in the German original is:

> Darum ist 'der Regel folgen' eine Praxis. Und der Regel zu folgen *glauben* ist nicht: der Regel folgen. Und darum kann man nicht der Regel 'privatim' folgen, weil sonst der Regel zu folgen glauben dasselbe wäre, wie der Regel folgen. (Wittgenstein, 2003b, § 202)

[11] We may call this the *Cartesian notion of thought*, comprising whatever one can be consciously aware of, i.e., everything which may enter one's stream of consciousness. For a

reasoning on the other hand. Reasoning and thinking are mutually exclusive while they are both kinds of (and together exhaust but do not exceed) conscious thought. The following tree diagram depicts how the three crucial notions relate to each other:[12]



It is certainly uncommon to draw a distinction between thought and thinking in this way, but it will be quite convenient to have a category where reasoning can be subsumed (thought) on the one hand and a category which reasoning can be contrasted with (thinking) on the other hand.

Let us first elaborate on the superordinate category: thought. The notion of thought I wish to make use of is akin to the definition of "thought" Descartes gives in his second replies to objections against his *Meditations*:

> I use the term *thought* to cover everything that is in us in such a way that we are immediately conscious of it. Thus all operations of the will, the intellect, the imagination, and the senses are thoughts. (Descartes, 2008, p. 102)

Descartes also explicitly tells us 'What thought is' in § 9 of part I in his *Principles of Philosophy*:

> By the word 'thought', I understand all those things which occur in us while we are conscious, insofar as the consciousness of them is in us. And so not only understanding, willing, and imagining, but also sensing, are here the same as thinking. (Descartes, 1982, p. 5)

compilation of eminent philosophers who also endorsed such an inclusive notion of thought – e.g., John Locke, George Berkeley, David Hume, Thomas Reid, Franz Brentano, Edmund Husserl, Immanuel Kant, and William James – see Bayne & Montague (2011b, pp. 4 f).

[12] I wish to thank Raquel Krempel and Césare Meurer at this point, who commented on a draft of my (2021) and thereby immensely helped me to set the terminology in this context straight. Possibly remaining faults are of course exclusively my own responsibility.

So, for Descartes, thought "[...] is any sort of conscious state or activity whatsoever [...]" (Williams, 2005, p. 62).

It is also important to note that the term "thought," as I use it here, exclusively covers *conscious* states or activities. The expression "conscious thought," which I used above,[13] is therefore pleonastic since "unconscious thought" – according to the terminology adopted here – amounts to a contradiction in terms. Friends of the Freudian tradition who like to talk about unconscious or subconscious thoughts might not be happy with this terminology. But the terminological choice I have made should not be overemphasized and does not imply any substantial disagreement in this regard in and of itself. This means that I do not dispute the existence of unconscious or subconscious cognitive states. I am merely not prepared to call cognitive states "thought(s)" if they are not conscious. Yet, no substantial conclusions about my view about the mental should be drawn from the fact that "thought" is here used in this technical sense.

The often noted ambiguity between "thought" as referring to a process or to having a thought on the one hand and as referring to a thought as the outcome of a process or the content of a mental occurrence on the other hand is unproblematic for this investigation. Context will usually make it sufficiently clear whether I am talking about one or the other when I use the expression "thought." In this section, however, "thought" is always meant to refer to a cognitive process, not to its content. The same holds true for my use of "thinking" and "reasoning," to which I shall turn now.

1.3.1 Dual Process Theories

It might be tempting, at first glance, to equate the distinction between thinking and reasoning with Daniel Kahneman's (2013) distinction between System 1 and System 2 processes. Kahneman's widely used distinction comes in some respects close to the differentiation I aim at: System 1 works automatically and without conscious effort for the cognizer, just like the kind of cognitive processes I wish to call "thinking" here. System 2 is slow and

[13] At the end of the penultimate sentence before the tree diagram.

effortful, often even arduous from a first-person perspective, just as reasoning is supposed to be. Further on, thinking and System 1 are both intuitive while reasoning and System 2 demand training and concentration.

Mercier & Sperber (2017, p. 46) present a convenient list of typical contrasting features for System 1 and System 2 processes:[14]

Type 1 processes	*Type 2 processes*
Fast	Slow
Effortless	Effortful
Parallel	Serial
Unconscious	Conscious
Automatic	Controlled
Associative	Rule-based
Contextualized	Decontextualized
Heuristic	Analytic
Intuitive	Reflective
Implicit	Explicit
Nonverbal	Linked to language
Independent of general intelligence	Linked to general intelligence
Independent of working memory	Involving working memory
Shared with nonhuman animals	Specifically human

It might therefore seem to be legitimate, at first glance, to identify thinking with processes carried out by System 1 whereas reasoning coincides with System 2 processes. Thought remains the all-encompassing category of conscious mental occurrences we started out with while thinking can be defined as as all kinds of thought which are not reasoning. However, some characteristics of System 1 processes are in conflict with my characterization of thinking. The most obvious issue is that Mercier & Sperber (2017) characterize System 1 processes as unconscious in the table on the current page.

[14] A quite similar table can also be found in Frankish & Evans (2009, p. 15). In comparison with the table printed on this page, Frankish & Evans (2009, p. 15) also add that System 1 is "evolutionary old" while System 2 is "evolutionary recent." Otherwise, apparent differences rather concern mere terminological variations. See also the diagram in Mercier & Sperber (2017, p. 65).

Since thinking is a subclass of thought and all of thought is conscious, if all System 1 processes are unconscious, then no System 1 process is correctly identified with thinking. There is no need to decide how central the characteristic of being unconscious is for System 1 processes right now. But if being unconscious should turn out to be a defining feature of System 1 processes, then thinking is crucially different from System 1 processes. However, Kahneman's distinction and my differentiation between thinking and reasoning definitely come apart in at least one further central respect.

The crucial difference between thinking and reasoning is that reasoning, in contrast to thinking, is subject to a certain kind of *correctness conditions*. These correctness conditions can be thought of in analogy to the conditions an argument needs to fulfill in order to be valid. The analogy between validity and the kind of correctness conditions of interest at this point is crucial. Truth conditions are, of course, also a kind of correctness conditions. Yet, thinking of correctness conditions for reasoning in terms of truth conditions instead of validity conditions would be a mistake. I will mean validity conditions whenever I use the expression "correctness conditions" in what follows, unless explicitly indicated otherwise.

The applicability of correctness conditions (in the relevant sense just fixed) is orthogonal to Kahneman's distinction because, e.g., causal reasoning – which is done by System 1 – has correctness conditions just like mathematical, statistical, or logical reasoning, which fall under the domain of System 2. The notion of reasoning which is of interest here may be called "System ≥ 1.5 reasoning" (McHugh & Way, 2018, p. 193) or "reasoning that is System 1.5 and up" (Boghossian, 2014, p. 2), i.e., reasoning that is person-level, conscious, voluntary, and active (cf. Boghossian, 2014, pp. 2 f; as well as McHugh & Way, 2018, p. 168), and – pace Boghossian (cf. 2014, p. 3) – often even effortful and demanding. Furthermore, reasoning is topic-neutral and domain-general (cf. Hurley & Nudds, 2006, p. 11).

The relevant notion of reasoning therefore comprises all System 2 processes and certain System 1 processes,[15] namely those which have correctness

[15] I ignore the question whether being unconscious is a defining characteristic of System 1 processes for the moment, but I tend to think that being unconscious is rather a typical

conditions. An example for a System 1 process which counts as reasoning is, as mentioned already above, causal reasoning, given that it proceeds consciously. The thought process which is, e.g., at play in a free word association test in the tradition of Carl Gustav Jung is an example for a System 1 process which does not count as reasoning. Despite the fact that there probably exist (more or less) normal and abnormal associations, there are no correct or incorrect associations in this case. The thought processes in question therefore count as thinking, not as reasoning, because they are not subject to correctness conditions in the way reasoning is.

1.3.2 Reasoning

The characteristics listed on the preceding page are not supposed to be a definition of "reasoning," but the properties used to characterize reasoning still need some elaboration. Reasoning happens on the personal level and is not a sub-personal process. Even though there are also sub-personal processes involved in reasoning, reasoning is a conscious act. That reasoning is a voluntary and active process also means that it is under the control of the reasoner. Reasoning is not something which somehow "happens" to a reasoner or what a reasoner suddenly and somehow finds herself doing. Reasoning is voluntarily initiated and actively controlled by the reasoner. Although there certainly are sub-personal processes involved in reasoning, the process of reasoning itself is fully conscious.

We may illustrate this last point by analogy with feeling a pain or a tickle, for example. While it takes a plethora of sub-personal and unconscious processes to create a feeling, feeling something is still an entirely conscious mental phenomenon. It does not make sense to say that someone is feeling something but is not aware of his feeling. If you are not aware of (feeling) any pain, then there simply is no pain even if your C-fibers are firing.[16] Although

and not a necessary characteristic of System 1 processes. This interpretation is compatible with the view that at least some System 1 processes are also conscious, so Kahneman's distinction can serve as a useful canvas to clarify the distinction between thinking and reasoning.

[16] I use the expression "C-fiber firing" as a placeholder for whatever neurophysiological or physical process corresponds to pain, as is commonly done in philosophical discussions

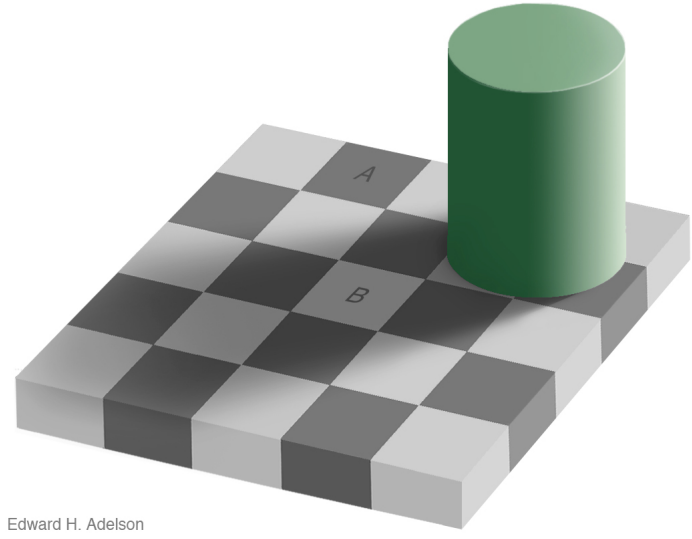


Figure 1.1: The *Checkershadow Illusion* by Edward H. Adelson

we know that feeling something is not a merely passive process of absorbing tactile or other bodily information, it is still not active and voluntary in the way reasoning is. I cannot decide what I feel in the way I decide what I reason about. It might also sometimes happen that I get "carried away" by my thoughts, that I merely float on my stream of consciousness, so to say. But reasoning is active and voluntary in a way which does not permit this kind of "mental drifting." The process of reasoning is controlled by the reasoner.

We should also note that being an active process and being a voluntary process come apart. Seeing something, i.e., visual processing, is now widely accepted to be a quite active process, far from merely taking in visual information. The fact that vision is an active process can be illustrated by Edward H. Adelson's well-known *Checkershadow Illusion*. We perceive square A as evidently darker than square B in figure 1.1. In fact, both squares are printed in exactly the same shade of gray, as can be seen in figure 1.2 on the next page.[17] The apparent brighter color of square B is due to the context in the

of this topic.

[17] The images displayed as figure 1.1 and figure 1.2 are available for free use and distribution on `http://persci.mit.edu/gallery/checkershadow`.

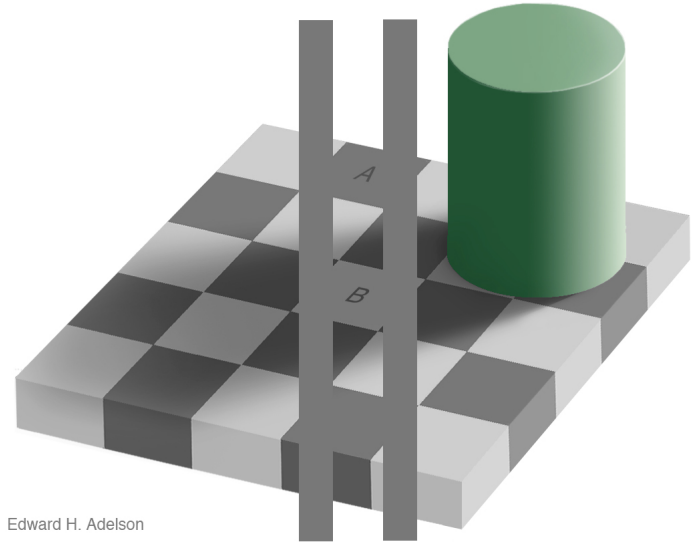


Figure 1.2: Proof that A and B are of the same color

depicted scene which presents square B, but not square A, as being in the shadow of the green cylinder. Despite the fact that square A and square B are printed in exactly the same color, we perceive square A as being darker than square B. This makes sense, since were we presented with an actual physical, three-dimensional scene which corresponds to the picture, square A would in fact be darker than square B although they reflect the same amount of light. Several entirely unconscious cognitive mechanisms make us perceive square A as being darker than square B even though this is not the information which reaches our eyes before unconscious cognitive processing interferes. If vision was a merely passive process of taking in the visual information that reaches our eyes, we would see square A and square B as having the same color since the same amount of light reaches our eyes from the relevant positions in the picture, i.e., the squares marked with *"A"* and *"B"* respectively. A plethora of (entirely unconscious) cognitive processes is necessary to perceive the illustrated scene as we do, where the lighting conditions are compensated, instead of merely perceiving patches of color which correspond to the actual color of the ink used to display the picture.

An analogous point can conveniently be illustrated with figure 1.3, an

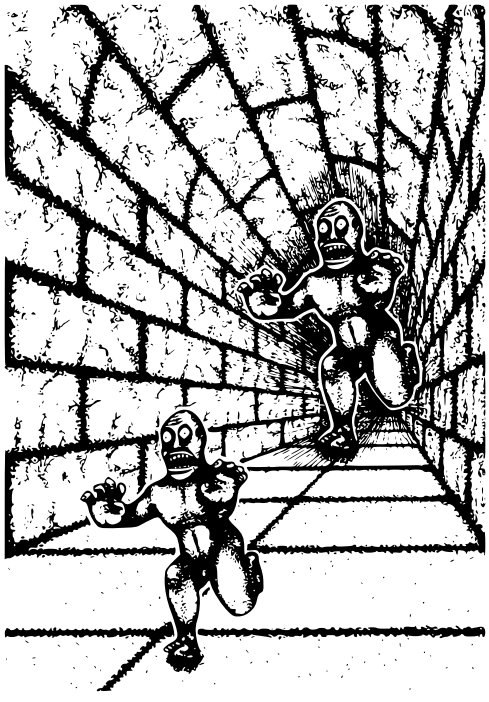


Figure 1.3: *Terror Subterra* (Shepard, 1990, p. 47)

illustration by Roger N. Shepard, where the monster in the back (monster 2) appears to be almost twice as large as the monster in the front (monster 1). Monster 1 and monster 2, however, take up exactly the same amount of space on the paper. That is to say, they have exactly the same size. The "illusion" that monster 2 appears bigger than monster 1 is due to sub-personal cognitive operations of our system of vision again. Monster 2 is depicted as being much further away down the tunnel. In a real situation where one object is more distant than another while the projections of both objects take up the same amount of space on our retina, the more distant object *is* much larger than the closer one.

So, seeing is evidently an active process. Yet seeing is not under voluntary control as reasoning is since I cannot decide what I (wish to) see in the way I can decide what (I wish) to reason about. Both characteristics, being active and being voluntary, are necessary for reasoning. While many instances of thinking are active, e.g., feeling or seeing something, they are not voluntary

and self-governed actions like reasoning. Reasoning is not only active but also voluntary since it is "[...] something we do, as opposed to merely being a causal process that takes place in our minds." (Boghossian, 2016, p. 3392) Although many cognitive processes are active, as in the case of vision, they are still not voluntary in the way reasoning is.

Another point worth emphasizing is that any train of thought, however loosely connected or wildly associative it might be, counts as thinking whereas reasoning crucially turns on the (type of) *transition* between mental contents. We may speak of reasoning only if those transitions are rule-guided. This restriction, of course, does not preclude flawed or incorrect reasoning. We need to make room for the possibility of (correctly) following the wrong rules as well as that of incorrectly following (appropriate) rules in reasoning. It just needs to be possible to reflect on any given transition in reasoning to double-check whether a given rule was followed correctly and whether the applied rule is appropriate for the respective situation and purpose at hand. To put it in terms of Wittgenstein's important distinction: We can only speak of reasoning if a cognizer is *following a rule* while thinking may proceed either merely *according to a rule* or even chaotically in a way that does not reveal any consistent pattern of mental transition at all.

1.3.3 Terminological Comparison

It is prudent to contrast the terminology adopted here with uses in other fields. This should reduce the risk of confusion as much as possible. Psychology and cognitive science especially deserve consideration since thinking and reasoning arguably fall in their domain. An authoritative, though tentative, characterization of "thinking" from the *Oxford* as well as the *Cambridge Handbook of Thinking and Reasoning* is the following:

> *Thinking is the systematic transformation of mental representations of knowledge to characterize actual or possible states of the world, often in service of goals.* [...] To count as thinking, the [transformations] must be *systematic* transformations governed by certain constraints. Whether a logical deduction or a creative

> leap, what we mean by thinking is more than unconstrained associations [...]. Often (not always – the daydreamer, and indeed the night dreamer, are also thinkers), thinking is directed toward achieving some desired state of affairs, some goal that motivates the thinker to perform mental work. [...] Thinking often seems to be a conscious activity of which the thinker is aware (*cogito, ergo sum*); however, [...] some mental activities seem pretty much like thinking, except for being implicit rather than explicit [...]. (Holyoak & Morrison, 2005, p. 2)[18]

The characterization of thinking in this quote cuts across the notions of thinking and reasoning adopted here. Although the characterization of thinking from Holyoak & Morrison (2005; 2013) arguably aims at capturing the most general term for mental activities, it also does not coincide with the Cartesian notion of thought I use as an umbrella term for all conscious mental occurrences. Carving out the individual differences can shed additional light on the terminology I use in this investigation.

Holyoak & Morrison's (2005; 2013) characterization of thinking is, at the same time, narrower and wider than the notion of thinking I use. Holyoak & Morrison's (2005; 2013) notion is narrower because they emphasize the importance of "[...] *systematic* transformations governed by certain constraints" for thinking. This restriction does not apply to my notion of thinking. The same holds for their claims that "[...] thinking is more than unconstrained associations [...]" and that thinking is directed towards a goal. "Thinking," as I use the term, covers especially lines of thought which do not conform to these restrictions. Systematicity and goal-directedness better fit my notion of reasoning. Holyoak & Morrison (2005; 2013) also do not wish to restrict their notion of thinking to *conscious* mental states. What they would call "implicit mental activities" or simply "implicit thinking" might, apparently, be sub-personal cognitive states, i.e., cognitive states which are unconscious or at least subconscious. This makes their notion of thinking wider than my

[18] Holyoak & Morrison (2013) is virtually identical with Holyoak & Morrison (2005) in the relevant respects. The passage quoted above corresponds to Holyoak & Morrison (2013, pp. 1 f).

notion of thinking and even wider than my notion of thought. Everything which is called "thought" here is – based on the Cartesian tradition – *essentially* conscious. Since thinking and reasoning are subcategories of thought, they are of course entirely conscious as well. Holyoak & Morrison's (2005; 2013) remark that daydreaming as well as ordinary dreaming while asleep are both thinking conforms quite well with my notion of thinking.

What Holyoak & Morrison (2005; 2013, p. 2) call "reasoning" is also narrower than the notion of reasoning employed here. They contrast reasoning with, e.g., judgment and decision making, as well as problem solving (cf. Holyoak & Morrison, 2005; 2013, p. 2) while the notion of reasoning which is being employed here incorporates all of these. Holyoak & Morrison (2005; 2013, p. 2) relate *reasoning* closely with logic, *judgment* seems primarily concerned with moral and evaluative assessment, *decision making* is active when we choose among alternatives, and *problem solving* in their presentation seems to roughly amount to what is often called "practical reasoning." My notion of reasoning comprises all of these aspects. This does not mean that I think that judgment, decision making, and problem solving are one and the same thing. The defining characteristic of reasoning I build on is simply general enough to include all of these different mental activities, so that any additional discrimination in the field of reasoning is unnecessary for the investigation at hand.

For a quick summary of the relevant notions, as they are used for the purpose of this investigation, we can say that

Thought is any conscious mental state or activity whatsoever and comprises thinking as well as reasoning.

Reasoning is any line of thought which is subject to correctness conditions (in analogy with validity conditions, not truth conditions).

Thinking is any line of thought which does not qualify as reasoning. This includes isolated beliefs although they are subject to a certain kind of correctness conditions. Yet the correctness conditions of beliefs correspond to truth conditions, not to validity conditions.

A clear understanding of these three notions is crucial, since the corresponding expressions are used in a clearly technical sense here. As depicted on page 12, "thought" names the supercategory which subdivides into thinking and reasoning. Thinking and reasoning do not overlap and together exhaust the whole domain of thought, *viz.*, thinking and reasoning are all kinds of thought there are as far as the present investigation is concerned. Although finer divisions can, of course, legitimately be drawn, this coarse-grained differentiation is adequate for my aims in this investigation. A more sophisticated discrimination would merely make the following discussion more complicated without adding any value (such as accuracy) to the discussion at hand.

Apart from the three central notions listed above, I frequently use the terms "cognitive state" and "mental state." I do not strictly distinguish between these terms and usually use them interchangeably. Importantly, while every thought is a cognitive/mental state, not every cognitive/mental state is a thought since cognitive/mental states can be either conscious or unconscious. The notion of a cognitive or mental state is therefore considerably wider than my notion of thought (which is the widest notion of *conscious* cognitive or mental states).

Chapter 2

No Private Reasoning

2.1 The *Private Reasoning Argument* (*PRA*)

In a next step, after the distinction between thinking and reasoning and the argumentative structure of *PLA* are sufficiently clear, we can now apply what we have so far to the domain of language and the mind – or, more precisely, to reasoning – by presenting what I call the *Private Reasoning Argument* (*PRA*). Before presenting the actual argument, however, a few words of explanation are in order.

Speaking about the structure of *PRA* in very general terms, we can say that the argument relies on two crucial steps: The first consists in the claim that reasoning presupposes publicly available correctness conditions and that a reasoner needs the ability to have access to these correctness conditions. This first step of the argument builds up until premise $R\,4$ below and is motivated by the rule-following constraint, which is also the crucial assumption behind *PLA*. The second decisive step finds expression in premise $R\,5$, which claims that the requirement from the first crucial step can only be accomplished with language. In an even more simplified manner we can say that reasoning requires (access to) publicly available correctness conditions (first step) and that (access to) publicly available correctness conditions requires language (second step), so reasoning requires language.[1] So much for the

[1] I am indebted to Frank Hofmann for very help- and insightful discussions about the

general underlying argumentative strategy. Here is the argument:

***PRA*:** The *Private Reasoning Argument* [2]

$R1$ * Reasoning is a rule-guided mental activity which consists (*inter alia*) in following inference rules.[3]

$R2$ * Reasoning is only possible for someone who can access and apply inference rules.[4]

$R3$ * Inference rules can only be accessed and applied by someone who can draw a distinction between correct and incorrect applications of inference rules.[5]

$R4$ * A distinction between correct and incorrect applications of inference rules can only be drawn by someone who has access to publicly available correctness conditions.

$R5$ * Only language (*viz.*, public language) is fit to provide access to publicly available correctness conditions.[6]

$R6$ Only (being in possession of) language allows drawing the distinction between correct and incorrect applications of inference rules. [via hypothetical syllogism from $R4$ and $R5$]

general structure of the Private Reasoning Argument.

[2] The asterisk (*) marks assumptions again.

[3] As McHugh & Way (2018, p. 184) would specify "[i]n a slogan, reasoning is rule-following that aims at fittingness." For the purpose at hand, however, it is only important that rule-following is *necessary* for reasoning. Whatever characteristic besides rule-following may be needed to be sufficient for reasoning is not relevant in this context.

[4] It is probably fair to ask whether premise $R2$ really represents an additional assumption or whether this claim is already entailed by premise $R1$. In any case, I prefer to make this step explicit even if it might be contained in premise $R1$ already if only to make the connection between reasoning and *reasoner* better visible.

[5] The ability to apply inference rules and distinguish correct from incorrect applications of an inference rule need not be flawless, but at least a general awareness of the fact that rules can be applied correctly or incorrectly is required for someone to count as a reasoner. Premise $R3$ might appear to be considerably stronger than it actually is since the ability to draw a distinction between correct and incorrect applications of a rule is in fact not overly demanding, as will be discussed in more detail in section 3.3.

[6] Or equivalently: Publicly available correctness conditions (for inference) can only be accessed linguistically.

$R\,7$ Only language provides access to inference rules. [via hypothetical syllogism from $R\,3$ and $R\,6$]

$R\,8$ $\therefore$ Reasoning is only possible for someone who is in possession of a language. [via hypothetical syllogism from $R\,2$ and $R\,7$]

$R\,9$ $\therefore$ Reasoning is not available without (being in possession of a) language. [logically equivalent with $R\,8$]

This argument roughly mirrors the structure of *PLA*. Premises $R\,1$ and $R\,2$ should not be too controversial, just like the first two premises of *PLA*, i.e., $L\,1$ and $L\,2$. In order to strengthen premise $R\,2$, we can nevertheless point out that, without the requirement that inference rules need to be accessible for a reasoner, we could not properly distinguish between actual reasoning and merely having an intuition. We do not need to engage in reasoning in order to have an intuition although we need reasoning to justify or question an intuition.

Premise $R\,3$ of *PRA* corresponds to *PLA*'s $L\,3$ while the commitments of *PLA*'s central premise $L\,5$ are roughly split up into two premises in *PRA*, namely premises $R\,4$ and $R\,5$. This is in part due to the fact that *PRA*'s $R\,5$ presupposes the conclusion of *PLA* and is therefore parasitic on the Private Language Argument. This is why *PRA* can only be successful if *PLA* (or another argument with the conclusion that a private language cannot exist) is acceptable as well. $R\,5$ of *PRA* is also crucial in another respect since it asserts that nothing but language could provide access to publicly accessible correctness conditions (for inference rules). This crucial assumption demands independent backup, which will be provided in the following section 2.2.

Before we come to this, note one striking dissimilarity between *PLA* and *PRA*: While *PLA* leads to the conclusion that a private language cannot exist, *PRA* does not conclude in saying that there cannot be private reasoning. This conclusion, however, can be easily derived from the characterization of reasoning in $R\,1$ and the basic Wittgensteinian insight that we cannot follow a rule in private (i.e., the *rule-following constraint*), which fuels *PLA* as well as *PRA*. With this presumption at hand, a derivation of the impossibility of private reasoning is trivial: If we cannot follow a rule in private, we also

cannot reason in private since reasoning is following inference rules.

2.2 Why *Only* Language?

The argument for the claim that *only* language possession enables reasoning (by providing access to publicly available correctness conditions, as $R\,5$ has it) is, in a nutshell, that no other plausible candidate is available to do the job. In order to provide access to correctness conditions/inference rules, a medium must be fine-grained enough to allow a precise formulation of the rules/conditions in question. In order to do so, sensitivity to intensional contexts *and* the availability of intersubjectively accessible communication are necessary.

Availability for intersubjectively accessible communication is necessary because it guarantees the possibility of reliably double-checking a given line of reasoning for correctness. If it was not even in principle possible to share a given line of reasoning, it is hard to see in which way we could be justified in calling a line of reasoning correct or incorrect. According to which standards? My own, *private* standards of correctness? Can that amount to anything more than a mere intuition of correctness? If not, we are back in a situation where everything that seems correct is correct, which means that any alleged distinction between correctness and incorrectness breaks down.

Being sensitive to intensional contexts is required because lines of reasoning (which may be expressed in arguments) are sensitive to intensional contexts as well. Therefore, no medium which does not exhibit this sensitivity could possibly be fit to take the role we have reserved for language, i.e., providing access to correctness conditions. Since we are concerned with reasoning, the requirement is probably not only sensitivity to intensional contexts but sensitivity to *hyper*intensional contexts. In contrast to intensional contexts, where co-extensional or extensionally equivalent representations cannot be substituted *salva veritate*, not even necessarily co-extensional representations can be substituted *salva veritate* in hyperintensional contexts.[7] Given that the correctness of reasoning turns on a relation

[7] It is more common to speak of co-extensional or extensionally equivalent *expressions*,

(or relations) which can be expressed with words like "because," "therefore," "since," and similar expressions and given that these expressions do not only create intensional but hyperintensional contexts (cf. Berto & Nolan, 2021, §§ 1.1.1 and 1.2), any medium which is fine-grained enough to capture the rules/conditions in question probably needs to be sensitive to hyperintensional contexts.

I will, however, ignore this stronger requirement (not only intensionality, but actually hyperintensionality) in what follows because the relevant points can be demonstrated by recourse to intensionality already. Despite the fact that I gesture towards an argument which builds on hyperintensionality to rule out images, maps, and diagrams as candidates to provide access to correctness conditions for reasoning on page 46, hyperintensionality is not needed to foster the claims I wish to make in the following sections. So, additional complications which would be dragged into the discussion by building on necessary sensitivity to hyperintensionality, instead of mere intensionality, can and therefore will be avoided.

2.2.1 The External World

The requirement of necessarily being sensitive to intensional contexts especially excludes reality itself as providing us with access to correctness conditions for reasoning. While reality is certainly intersubjectively accessible, it is not sensitive to, e.g., different descriptions of one and the same thing or state of affair. Reality is indifferent to my depicting one and the same person as either 'the 45$^{\text{th}}$ president of the United States' or 'the worst president of the United States ever.' In reasoning, however, this difference can be crucial, since

> He will not be reelected because he is the 45$^{\text{th}}$ president of the United States.

is not an inference of the same quality and plausibility as

but, since I do not wish to beg the question regarding *linguistic* representations, I prefer to speak of representations in general in this context.

> He will not be reelected because he is the worst president of the United States ever.[8]

These kinds of differences must be captured by a medium which can provide access to correctness conditions for reasoning. The external world – *viz.,* reality – is, however, insensitive to intensionality in the relevant respect. It probably does not even make sense to call reality "intensional" or "extensional" since these properties seem to be applicable to representational systems only. Saying that reality or the external world is (sensitive to) intensional (contexts) most likely amounts to a category mistake. However, insofar as nothing could possibly provide access to correctness conditions for reasoning without being (sensitive to) intensional (contexts), reality is ruled out, regardless of whether the statement "reality is intensional" is false or nonsensical.

One way to make sense of the idea that reality might be (sensitive to) intensional (contexts) is to make use of a *Lagadonian language* (cf. Lewis, 1986, pp. 145 f).[9] In a Lagadonian language (with a Lagadonian interpretation) everything simply serves as a sign of itself.[10] This way we can regard reality as a representational system, and it seems entirely legitimate to ask

[8] This line of reasoning is, of course, highly enthymematic and presupposes an implicit premise, according to which bad presidents do not get reelected. Also, these lines were formulated in 2019 to refer to the now former president. However, an adaptation in the tense of the formulation above should be unproblematic, if needed.

[9] The name "Lagadonian language" comes from *Gulliver's Travels*, but it is not the language actually spoken in Lagado, and it is never referred to as "Lagadonian" by Swift (2019). Lagado is the metropolis of the kingdom of Balnibari where the Grand Academy of Lagado is situated. What is called "Lagadonian language" here is inspired by a linguistic project, adopted by "[...] many of the most learned and wise [...]" (Swift, 2019, p. 149) scholars of this institution. The vast majority of the inhabitants of Lagado speaks the common language of Balnibari, which probably coincides with the language spoken in Laputa, the flying island from where Balnibari is governed – or rather dominated. This language is "[...] not unlike in sound to the Italian [...]" (Swift, 2019, p. 127), so it is obviously an ordinary *spoken* language. Swift presents the language project which is called "Lagadonian language" here, despite the fact that it is not the language of Lagado, towards the end of part 3, chapter 5 of *Gulliver's Travels* (cf. Swift, 2019, pp. 149 f).

[10] A Lagadonian language, like every representational system, can have different interpretations (cf. Lewis, 1986, p. 146). Several different Lagadonian languages are therefore possible. Since the only version of interest here is a Lagadonian language where everything stands for itself (i.e., the Lagadonian interpretation), I will exclusively talk about *the* Lagadonian language, instead of *a* Lagadonian language, from now on.

whether the Lagadonian language (*qua* representational system) is sensitive to intensionality or not. But the Lagadonian language is purely extensional, so this move does not help with getting access to (inferential) correctness conditions for reasoning. Still we can, at least, give the question of whether reality is (sensitive to) intensional (contexts) a sensible reading.[11]

This also proves that communicative attempts such as pointing to, e.g., things or events in order to fix an inference relation between certain occurrences pointed to is too coarse-grained for reasoning.[12] This is because different descriptions of what was pointed to can affect the quality of the reasoning involved, as can be readily seen from the example on page 29. We therefore need a more sophisticated medium which allows us to have access to correctness conditions for reasoning. The most plausible candidate to accomplish this, apart from language and reality, is surely some kind of mental representation.

Before we come to that and leave the external world behind, however, we must note that a central claim I made before might be disputed, namely the claim that the external world is not a representational system. Braddon-Mitchell & Jackson (2007, p. 181) clearly disagree:

> We can be certain that something map-like can serve to represent any empirical fact about our world. The world itself is map-like: it is a vast array in space-time, rather than a two-dimensional configuration on paper, but that difference is inessential to its map-like status. And, of course, the world itself makes true each and every fact about our world; it is a *perfect* representation of itself.

If the world is a perfect representation of itself, then not only drawing on

[11] I wish to thank Frank Hofmann for pointing me to David Lewis's discussion of the Lagadonian language.

[12] This also holds true for, e.g., the so-called *hands & feet system of communication*, mentioned in Lohmar (2016, pp. 169 f), where we try to make ourselves understood without a common language by making use of gestures and pantomime. The hands & feet system of communication, however, must not be confused with sign languages. The latter are codified systems of communication which count as languages whereas the former is merely an intuitive and non-standardized means of expression which does not count as a language.

the Lagadonian language in order to answer the question of whether reality might be sensitive to intensionality is superfluous, but also my claim that the external world is insensitive to intensional contexts must be mistaken "[...] if the world itself makes true each and every fact about our world" (Braddon-Mitchell & Jackson, 2007, p. 181). But in the quote above Braddon-Mitchell & Jackson, interestingly, silently switch from talking about "[...] any *empirical* fact about our world" (Braddon-Mitchell & Jackson, 2007, p. 181; emphasis added) to talking about "[...] *each and every* fact about our world" (Braddon-Mitchell & Jackson, 2007, p. 181; emphasis added). This does not seem to be an unproblematic transition.

Even if we were to admit that the world represents itself and that it represents every empirical fact, why should we think that each and every fact is an empirical fact?[13] There are also logical and mathematical facts, e.g., facts about transfinite numbers, which can hardly count as empirical facts. How could such facts be represented in or by the empirical reality? Even if the world might represent every empirical fact, we arguably need a medium which is capable of representing not only empirical facts but also facts which go beyond the empirical. An array in space-time, however vast it may be, is probably not capable of doing so.

I therefore think that we should, pace Braddon-Mitchell & Jackson's objection, stick to the claim that the external world is – even if we were to consider it as a representational system, which I think we should not – too coarse-grained to provide access to correctness conditions for reasoning. I also tend to agree with Camp (2007, p. 179, n. 32) that "[...] many empirical facts depend upon counterfactual relations, which even a 'vast array in space-time' as large as the world itself doesn't suffice to represent."[14] Con-

[13] See also Camp (2007, p. 179, n. 32) for a similar point that Braddon-Mitchell & Jackson (2007) fail to provide "[...] an additional closure condition to the effect that the world is all that is the case."

[14] Exactly the same passage from Braddon-Mitchell & Jackson (2007) I quoted on the previous page is also provided and discussed by Camp (2007). She even claims that "[...] in important respects the entire world is *less* expressively powerful than an ordinary [...] road map, because it doesn't have symbolic icons like 'Philadelphia' written on it." (Camp, 2007, p. 179, n. 32) I find this point less convincing than other considerations Camp (2007) has to offer, but I definitely agree with Camp's skeptical reaction towards Braddon-Mitchell & Jackson's claim. However, I will return to maps and Camp's (2007)

sequentially, the world might even fail to represent every empirical fact, not to speak of non-empirical facts.

However, even if we were to consider the world as a representational system, and even if the world was a sufficiently fine-grained representational system, and even if the world represented not only empirical but also abstract facts, it still seems that the world could not represent all these facts in a way which would make them suitably available for intersubjectively accessible communication. In order to make the topic of my reasoning available for intersubjectively accessible communication, I need to present said topic in an intersubjectively accessible way, for example by pointing to it. Even if the world itself is sufficiently fine-grained for the purposes of reasoning, my "pointing" is probably not fine-grained enough. How could I point out, let us say, the smallest prime number instead of the even prime number in a way which makes it intersubjectively clear that I mean the one but not the other?[15] This certainly does not make the external world look like an attractive candidate for providing correctness conditions for reasoning. Let us face some more promising candidates.

2.2.2 Mental Representations and Language of Thought

Mental representations are without any doubt sensitive to intensionality (and undisputedly even to hyperintensionality; see pp. 29 f), but the trouble with mental representations (even if this does not hold for their contents; see foot-

considerations in this regard in section 2.2.3.

[15] That this example builds on an abstract object and a hyperintensional difference exacerbates the problem, but it is not essential to make the relevant point. The question is: How can I point to the 45th president of the Unites States/the worst president of the United states ever in a way which makes it intersubjectively clear under which guise I mean to point out this concrete object without the aid of a communicative medium? (We might also reformulate the question to ask how I can point to an *intentional* object. How do you point to the Devil or Pegasus, or to the 45th president of the Unites States instead of the worst president of the United states ever?) We also know that the difference between rigid and non-rigid designation can affect reasoning. Can I non-rigidly point to one and the same object in different ways, or will my pointing always rigidly fix the object pointed to? (In the case of the smallest prime number and the even prime number I even had to rigidly point to the same object in different ways on top of the problem of how to point to an abstract object at all.)

note 25 on page 47) is that they do not seem to be publicly available – *viz.*, they are private – and therefore cannot accommodate the rule-following constraint.[16] Although it is clear that mental representation is necessary for reasoning, no combination of mental representations or transition from one representation to another, as long as they are unaided by language, could constitute reasoning. This is because, besides intensionality, the availability of intersubjectively accessible communication is needed as well. Mental representations as private entities are therefore unable to provide access to correctness conditions, which is required for reasoning.

This holds true even if we assume that mental representation, and thereby our vehicle for thought, is already language-like from the outset. Let us consider *Mentalese*, the alleged "language" of thought:[17] If Mentalese is not a public language – as Jerry Fodor (2008, p. 80) himself suggests – it does not even count as a possible language, following the conclusion of *PLA*.[18] If Mentalese (assuming that it actually exists) should turn out to be a public language, on the other hand, it can provide the required access to correctness conditions and inference rules, and the conclusion of *PRA* holds true because the so-called "language" of thought is a real, i.e., public, language.[19]

Regarding the language of thought hypothesis in general, i.e., the claim that we think in a language of thought (be it public and therefore an actual *language* of thought or private and therefore merely a "language" of thought), I wish to stay explicitly neutral. The account defended here is certainly compatible with the theory that there is a language of thought. However, a language of thought cannot be presupposed in an investigation regarding the relation between thought and language, as already indicated in footnote 17 on the current page. If all our thought is carried out in a language of thought,

[16] Mind premises $R\,3$ and $R\,4$, or just substitute "private representations" for "a private language" in $L\,5$ to see that mental representations are ruled out by the rule-following constraint (see p. 11).

[17] Some people might feel the urge to object at this point that presupposing a language of thought would beg the question in an investigation of the relation between thought and language, but bear with me for the moment. I will come back to this topic shortly.

[18] This is, by the way, also the reason why I talk about the alleged "language" (in scare quotes!) of thought in this context.

[19] For considerations regarding the possibility that a person's language of thought just is the same as a person's public language, e.g., English, see Devitt (2006, pp. 148 ff).

then not only reasoning but also thinking is language-dependent (or at least "language"-dependent). So, assuming that there is a language of thought implies that thought is dependent on language. Although this is a very substantial presupposition to make regarding the relation between thought and language, it would still not render the claim defended in this investigation trivial.

Since I argue that reasoning requires a *public* language, *PRA* still contributes a substantial claim to a theory which builds on a private language of thought like Fodor's Mentalese. Assuming that a public language also serves as the language of thought – see footnote 19 on the facing page – would of course trivialize *PRA*'s conclusion. If all our thoughts are encoded in a public language of thought, it trivially follows that also reasoning – which is a proper subset of thought – is carried out in language. There might still remain some modal crumbs to argue for if we assume that there is a public language of thought: Even if thought is *de facto* linguistically organized, the question whether this is necessarily – or even essentially – the case remains open. Also whether all kinds of thought are necessarily (or essentially) linguistic would not be automatically answered by assuming that there is a public language of thought. Since I do not endorse the view that we think in a language of thought, much less in a public language of thought, I will happily leave such questions to proponents of the language of thought hypothesis.

Laurence BonJour (1991) brings up rather general considerations against "the view that thought is fundamentally a linguistic or symbolic process which employs a representational system at least strongly analogous to a natural language." (BonJour, 1991, p. 331) He claims that "[...] at least some of the elements of thought must be intrinsically meaningful or contentful, must have the particular content that they do simply by virtue of their intrinsic, non-relational character." (BonJour, 1991, pp. 345 f) Insofar as mental symbols, if they are language-like, are not intrinsically meaningful, the language of thought hypothesis must be mistaken, according to BonJour[20] – at least if the language of thought is to govern all of thought. Since I do not endorse the language of thought hypothesis, it does not need to be settled at this point

[20] The gist of BonJour's argument is condensed in BonJour (1991, p. 336).

how cogent BonJour's arguments against "[...] the linguistic or symbolic conception of thought [...]" (BonJour, 1991, p. 345) actually are. That said, a minor mistake of BonJour's should be corrected here: He expresses the potentially widely shared opinion that the linguistic conception of thought, i.e., the language of thought hypothesis, even "[...] has a good claim to be the *defining* thesis of the linguistic or analytic school of philosophy" (BonJour, 1991, p. 331; emphasis added).[21] Although the assumption that we think in a language of thought remains widely shared among analytic philosophers, I would dispute that it is the defining thesis of analytic philosophy since I definitely identify my account with this philosophical tradition without endorsing the language of thought hypothesis.

Be that as it may, the position that the language of thought is a national language, such as, e.g., English, French, or German, is certainly a minority view among adherents of the language of thought hypothesis. It is often built on introspective evidence that we think in the language we speak. This might frequently be the case; and it might even be the case that our thoughts are often vague until they are fully formulated. Yet, the existence of imagistic thought is hard to deny (see section 2.2.3.2), and the view that we think in a national language has a hard time explaining several quite common phenomena. One of these is the tip of the tongue phenomenon. If we think in the language we speak all along, how can it happen that we are often missing the right word(s) to appropriately express what we think? This and similar questions will come up again in chapter 10. For now, however, still other competitors for the role to be filled by language await evaluation.

2.2.3 Mental Maps and Imagistic Reasoning

Another popular account of how thought might be organized, if not in a language of thought, is the idea that we think in terms of mental maps.

[21] BonJour repeats this claim in an unpublished paper about 'Analytic Philosophy and the Nature of Thought' (available at `https://faculty.washington.edu/bonjour/Unpublished%20articles/UBCPAPER.html`), where he bases it on Michael Dummett. It might therefore be considered a mistaken self-understanding of the analytical tradition itself, rather than a mistake of BonJour's.

Elisabeth Camp (2007) provides an impressive demonstration of which kinds of logical operations can be represented via maps, including negation, disjunction, implication, and even – to a certain degree – intensionality and quantification. Camp (2007) thereby proves that the representational capacities of maps reach far beyond what we traditionally assumed maps would be capable of. But the decisive question here is: Are the representational capacities of (mental) maps sufficient to enable reasoning?

2.2.3.1 A First Look at Mental Maps

Before directly engaging with this question, I should first state that I do not see the account presented here to be in conflict with anything Camp (2007) says. This is probably in part due to the fact that Camp's (2007) focus primarily lies on arguing against the language of thought hypothesis. Since I am not a proponent of the language of thought hypothesis, Camp's (2007) arguments against this view are hardly problematic for the account defended here. However, saying that Camp (2007) argues against the language of thought hypothesis is a rather rough characterization of her line of argument. What Camp (2007) actually argues against is what she calls *'Strong-LOT,'* i.e., the strong language of thought hypothesis, which claims that "[. . .] thought requires a specifically sentential [and thereby linguistic] structure and semantics." (Camp, 2007, p. 152) By contrast, Camp does not set out to refute what she calls *'Weak-LOT,'* i.e.,

> the claim that thought requires a system of representational vehicles with *some* recurrent constituents that can be recombined according to *some* set of rules to produce representations of systematically related entire contents. (Camp, 2007, p. 152)

In other words, this is a representational system with combinatorial or compositional syntax and semantics (cf. Camp, 2007, p. 177, n. 13). What Camp (2007) argues against is the view that Strong-LOT follows from Weak-LOT (cf. Camp, 2007, p. 152). Camp claims that mental maps fit what is required by Weak-LOT, and mental maps therefore provide a proper vehicle for thought. Since maps do not exhibit sentential structure and are therefore

not languages, we have a representational medium which fits Weak-LOT but not Strong-LOT. Mental maps therefore provide a possibility of fulfilling our representational and cognitive needs without the need to commit ourselves to a *language* of thought.

Camp (2007) presents a complex web of elaborate considerations which eventually lead her to settle for what Camp (2007, p. 169) calls

> [...] *Sophisticated-LOT*: the representational vehicle which underwrites highly flexible thought about abstract, hierarchically-structured states of affairs is likely to be sentential in form. Because the distinctive power of human cognition seems to depend on our agility at representing and manipulating such contents, this gives us good reason to think that much of our own cognition, in contrast to that of other animals, takes place in language. However, this conclusion depends crucially upon the specific contents that humans think about and what they do with those contents, and not on general features of thought *per se*.

In other words, Camp (2007) claims that thought in general does not presuppose a *language* of thought, but, as a contingent matter of fact, human thought in its highly flexible sophistication most probably takes place in a language of thought.

Be that as it may. I am, first of all, not primarily concerned with the question of how the *vehicle* of thought is actually structured. Secondly, I am not primarily concerned with the question of how the vehicle of *thought* is structured. I am here primarily concerned with reasoning, not with the more general notion of thought; my concern is not primarily the vehicle for thought, or the vehicle of reasoning, but the cognitive preconditions for reasoning. I prefer to stay neutral regarding the thought of language hypothesis, as well as regarding any theory about the actual structure of the vehicles we use for thinking or reasoning. As I will explain in section 3.2, my account is compatible with the assumption that we (at least sometimes) reason without employing language, *viz.*, reasoning does not need to take place *in* language, while it still holds true that (public) language is a constitutive precondition

for reasoning.

The details of how this is supposed to work will be, as I said, delivered in section 3.2. For now, it should be sufficient to say that I am not committed to the claim that reasoning needs to proceed in language, so I tend to agree with Camp (2007) that at least sometimes we probably reason with (or in, or via) maps. The crucial question for me is, as already mentioned, not which vehicle we use for reasoning. The question is whether any other medium but language can provide access to correctness conditions for reasoning and thereby make reasoning possible in the first place. I will neglect the requirement that the representational system in question also needs to be able to make said correctness conditions publicly available. Maps are certainly a publicly available representational medium, and, even though it is not so clear that the same also holds true for *mental* maps, I will set this issue aside and grant mental maps the benefit of doubt in this regard. What then of the potential for fine-grainedness of (mental) maps? Are maps at least potentially fine-grained enough to capture and represent correctness conditions which are needed to differentiate correct from incorrect reasoning and thereby make reasoning possible in the first place?

2.2.3.2 Imagistic Reasoning

Instead of directly providing an answer to this question right away, I would like to move the discussion to a different potential contestant against the claim that language is necessary for reasoning. This will allow me to illustrate an issue which, I think, also applies to mental maps as a candidate (instead of language) for enabling the capacity to reason. What I wish to consider in the next few pages is reasoning by means of thought experiments. Simon Stevin's famous thought experiment concerning inclined planes, being an impressive example of this variety of reasoning, will serve our purposes perfectly.

Let us consider the setup of Stevin's thought experiment as depicted in figure 2.1 on the next page: We have "[...] a prism-like pair of inclined (frictionless) planes with linked weights such as a chain draped over it." (Brown, 2011, p. 3) Since we are operating in an idealized situation, we should

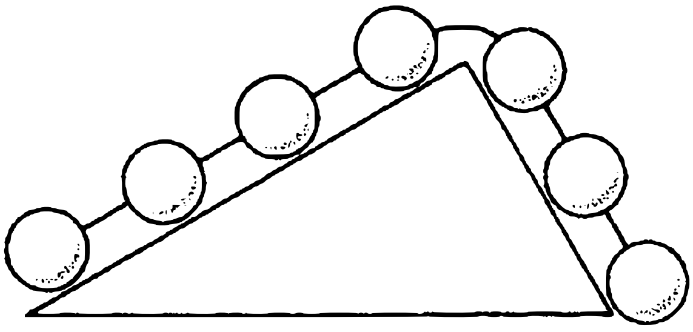


Figure 2.1: Stevin, the problem (Brown, 2011, p. 4)

also assume that every link in the chain is a perfect sphere with exactly the same mass so that the chain is absolutely uniform throughout. The question is now: What will happen with the chain in this idealized situation without any friction?

> There are three possibilities: It will remain at rest; it will move to the left, perhaps because there is more mass on that side; it will move to the right, perhaps because the slope is steeper on that side. (Brown, 2011, p. 3)

We can, of course, set out to calculate the forces on each side of the chain in order to determine the correct answer. Yet, there also exists a more elegant way to reach an answer to the question of what will happen with the chain in a situation that corresponds to the depiction in figure 2.1. If we take a look at figure 2.2 on the facing page, we see that we can find an answer to the question of what will happen to the chain from figure 2.1 by extending the chain to form a closed loop around the prism. Since the added chain parts are perfectly symmetrical, they exert equal force on both sides of the original chain from figure 2.1. The imagined modification we perform in figure 2.2 therefore does not change any aspect from the original situation of figure 2.1, as far as our question – namely, what will happen with the chain – is concerned. The chain in figure 2.1 will behave exactly like the chain in figure 2.2. But by thinking of the problem in terms of figure 2.2 – i.e., by considering how a closed loop will behave – the answer now seems obvious.

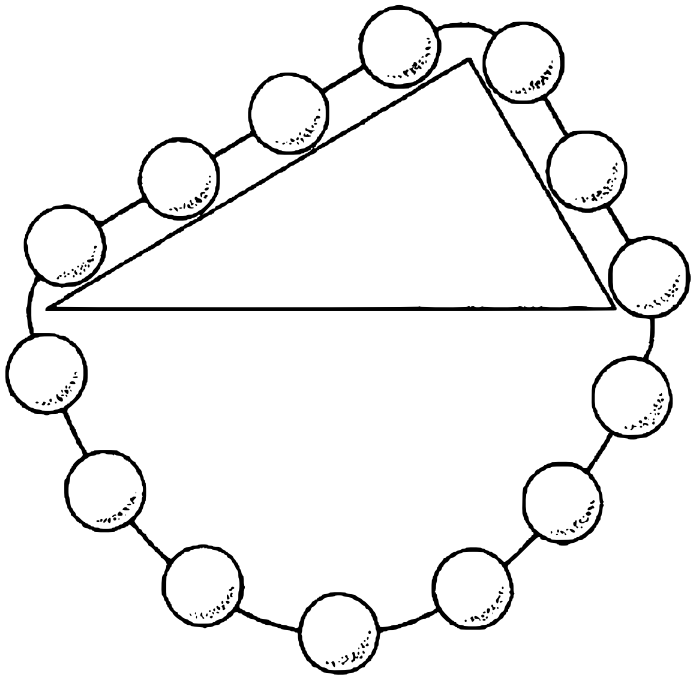


Figure 2.2: Stevin, the solution (Brown, 2011, p. 4)

While all three possibilities appear to be live options when we look at figure 2.1, we can clearly see that the chain will neither move to the left, nor to the right, when we look at figure 2.2. So, the correct answer is clearly that the chain in figure 2.2, as well as in figure 2.1, will remain at rest.

This is an impressive example not only of reasoning via thought experiments, but supposedly also of imagistic reasoning. We found the correct answer to the (I assume) initially puzzling question of what happens to the chain in figure 2.1 by imagining a scenario which is equivalent to figure 2.1 in every relevant respect but still allows us to readily grasp (and presumably also understand) the correct answer. This is certainly a quite elegant – not to say beautiful – specimen of reasoning. So, I am not in the slightest inclined to doubt, even for a second, that this is a genuine case of reasoning. What is especially relevant now is that it seems that language did not play any role in reasoning through the problem along Stevin's path, i.e., by making the transition from figure 2.1 to figure 2.2 in order to find the correct answer to the question of how the chain will behave.

It seems that language did not play the slightest role in reaching the correct conclusion, nor does language seem to be required to justify the conclusion that the chain will not move. I assume that the linguistic clarification of the thought experiment I provided (including the claim that we are operating in an idealized situation without friction) is dispensable. Likewise, coming to see which one of the three options coincides with the correct answer (roughly, context of discovery), as well as coming to see why the first option is the correct answer (roughly, context of justification), seem to be entirely independent of language.[22] So we should conclude that there are genuine cases of reasoning which do not depend on language, or so it seems. Yet I claim that this conclusion would be hasty since the imagistic reasoning which led us to the right conclusion alone cannot determine the correctness conditions for the line of reasoning which results in the correct conclusion.

In order to see why, we need to acknowledge that there are several lines of reasoning Stevin's thought experiment might give rise to:

1. We might come to the conclusion that the chain in figure 2.1 will remain motionless because chains like those in figure 2.2 do not tend to start rotating on their own.

2. Or we might come to the conclusion that the chain in figure 2.1 must remain motionless since, otherwise, the configuration in figure 2.2 would represent a perpetual motion machine, and, since perpetual motion machines are impossible, the chain must stay in place.

I think it is obvious that line of reasoning 1 is different from line of reasoning 2. But given that we might plausibly say that both lines of reasoning are valid and the goal is to carve out a relevant difference as clearly as possible,

[22] I will also assume that language is not needed to communicate the correct answer. We can assume that a testee is led through the thought experiment by being presented with a couple of video sequences which enact the three options so that the test subject can point to the correct solution when each sequence of the three options is played simultaneously (let us say, on three separate screens positioned next to each other) after the person was made to consider the situation in figure 2.2, or something along these lines. For a nice presentation of suitable video material for this case – at least if we cut out appropriate sequences and play them muted – see udiprod (2012) on `https://youtu.be/nDKGHGdXLEg`.

let us combine the lines of reasoning presented above (namely 1 and 2), to create a new line of reasoning 3, which is arguably invalid:

3. We might come to the conclusion that the chain in figure 2.1 *must* remain motionless because chains like those in figure 2.2 do not tend to start rotating on their own.

This line of reasoning (i.e., number 3) combines the conclusion from line of reasoning 2 with the explanation/justification/premise from line of reasoning 1. But given that we read "must" as indicating physical necessity throughout all lines of reasoning, the premise of reasonings 1 and 3 is too weak to warrant the conclusion of reasonings 2 and 3. Line of reasoning 3 is therefore invalid while lines of reasoning 1 and 2 are valid.[23]

The crucial point here is to note that imagistic reasoning, unaided by language, can lead us to identify the correct answer to the question of how the chain will behave. In order to count as reasoning, however, we said that it must be possible to check whether the line of reasoning was actually correct or incorrect, independently of the question of whether we found the right answer to the question. Remember that incorrect reasoning can result in true claims, just as invalid arguments can have true conclusions. While imagistic reasoning, unaided by language, can lead us to find the correct answer, this kind or reasoning is not sufficient on its own to establish genuine reasoning, which can be checked for correctness. This is, I claim, because imagistic

[23] I assume, for the sake of argument, that inductive generalizations – such as the one present in 1 and 3 – warrant factual conclusions, like the conclusion in 1, but not modal conclusions such as in 2 and 3. Line of reasoning 2, I assume, is valid because it draws a modal conclusion (namely that the chain *must* stay in place) from a modal premise (namely that a perpetual motion machine is *impossible*). Line of reasoning 1 is valid, I assume, because it draws a factual conclusion (namely that the chain *will* not move) on the basis of an inductive generalization (namely that certain things do not happen, based on previous experience). Line of reasoning 1 is therefore inductively valid. Line of reasoning 3, on the other hand, draws a modal conclusion (namely that the chain *must* remain motionless) from a factual premise (namely that a closed chain as depicted in figure 2.2 *does* not start to move on its own). As mentioned above, I presuppose that "must" is throughout all lines of reasoning interpreted as expressing physical necessity and is not, e.g., read in an epistemic sense – *viz.*, as expressing, for example, certainty instead of physical necessity. Given these preconditions, line of reasoning 3 is invalid because the conclusion that the chain *cannot* move does not follow from the claim that it *does* not move, which would be warranted by the inductive justification provided in 3.

reasoning, as long as it is unaided by language, cannot distinguish between lines of reasoning 1, 2, and 3. Each of these lines of reasoning, and certainly several other slightly different lines of reasoning, can be equally suggested by the imagistic basis of Stevin's thought experiment. Since pictures do not do a good job of distinguishing factual from modal claims, the purely imagistic basis of the thought experiment, as well as the line of reasoning if it is conducted purely imagistically, are too coarse-grained to distinguish between correct and incorrect lines of reasoning which correspond to valid and invalid arguments.

Imagistic resources alone, as long as they cannot be supplanted with the resources of language, are therefore insufficient for genuine reasoning, for which it must always be possible to double-check on its correctness. We can put the same point in different terms by making use of the aforementioned distinction between context of discovery and context of justification: Imagistic reasoning, unaided by language, can guide us to reach the correct conclusion, as far as the context of discovery is concerned. Imagistic reasoning, unaided by language, can also help us to find the right explanation (and thereby the correct justification) for the conclusion we reached. We have now left the context of discovery and entered the context of justification. So, I do not claim that imagistic reasoning cannot be useful in the context of justification. It can be quite useful also as far as justification is concerned, but not on its own. We now need to drop the qualification "unaided by language" when we enter the context of justification since only language provides the resources – *viz.*, only language is fine-grained enough – to distinguish correct lines of reasoning from incorrect lines of reasoning, which cannot be properly separated on an imagistic basis alone.

We will come back to matters of interest for this argument below in section 3.2. I will argue there that imagistic reasoning in the sense discussed here on the basis of Stevin's thought experiment, even to the extent that no language is involved – i.e., even imagistic reasoning unaided by language – is only available to linguistic creatures. This means that imagistic *reasoning*, even if it is unaided by language, can only be conducted by linguistic creatures. The claim that reasoning in general, regardless of its kind or medium,

is dependent on language can stay in place in the face of the existence of full-blown, genuine reasoning where no actual employment of language is involved. As I said, I will come back to elaborate on and explain this claim in section 3.2. But for now we should come back to our discussion of mental maps to see how this excursion into an obviously different topic – since mental maps do not seem to play any role in what we discussed in connection with Stevin's thought experiment – relates and contributes to the question of whether mental maps, instead of language, could enable reasoning.

2.2.3.3 Back to Mental Maps

The excursion into imagistic reasoning, exemplified by reasoning through Simon Stevin's thought experiment about inclined planes, is also relevant for our discussion about mental maps. Mental maps basically suffer from the same shortcoming which also prevents imagistic reasoning from functioning without language. Although Camp (2007) impressively demonstrates that maps have far greater expressive capacities than traditionally assumed, the expressive capacities of maps still fall significantly short of the expressive capacities needed for reasoning and provided by language.

Maps, be they mental or physical, will run into the same problem I have illustrated for images against the backdrop of Stevin's thought experiment: They are too coarse-grained – *viz.*, their representational capacities are insufficient – to make the often delicate distinctions which are needed to distinguish between *prima facie* similar but ultimately crucially different – due to differences in correctness/incorrectness of – lines of reasoning. If a medium is too coarse-grained to permit drawing a clear line between certain correct and certain incorrect lines of reasoning, then the medium in question will consequentially not allow a reliable double-check on the correctness or incorrectness of a given line of reasoning. To put it quite bluntly, a medium which confuses incorrect with correct lines of reasoning cannot provide access to the required correctness conditions to check for the correctness/incorrectness of a given line of reasoning. If it could provide this access, it would not confuse the correct line of reasoning with the incorrect line of reasoning in the first

place.

To be sure, I did not provide a conclusive argument for the claim that there cannot be even a single instance of non-linguistic reasoning which is not in need of corroboration by resources only language can provide. I think that this is actually the case since I claim that language is necessary for reasoning. But regarding reasoning with (mental) maps and imagistic reasoning, I have not conclusively demonstrated that this is actually the case. I have merely gestured towards this conclusion by making it plausible via illustrating how maps and images might fail to provide the fine-grainedness needed for reasoning. At best, this establishes the conclusion that not every kind of reasoning can be conducted without language. My claim is considerably stronger than this, namely that no kind of reasoning can be conducted without language.

It might be possible to formulate a conclusive argument to this effect on the basis of the claims that sensitivity to hyperintensional contexts is necessary for every kind of reasoning and that no other medium (including images and maps) could possibly be sensitive to hyperintensional contexts (and also be intersubjectively available, *viz.*, not private). This might be a promising line of argument, for the required premises indeed seem to be quite plausible: Since the expressions most intimately connected with inferential relations arguably create hyperintensional contexts – as mentioned on pp. 29 f – and since I would be very surprised to find a demonstration that other media than language – such as (mental) maps, (mental) images, or (mental) diagrams – are capable of drawing not only intensional [24] but even hyperintensional distinctions, we might conclude that an argument along these lines looks promising. I will leave matters at this rough sketch of a potentially prolific argument for the moment and keep the work necessary to transform this sketch into fleshed-out argumentation for a later occasion. For now, I must hope that my gesture towards the weaknesses of non-linguistic representational systems is sufficiently convincing for the inclined reader to, at least tentatively, accept the conclusion I endorse – namely that reasoning is constitutively dependent on public language – as far as the reasons I have provided for this claim reach.

[24] As demonstrated, to a certain degree, for maps in Camp (2007, pp. 164 ff).

2.2.4 Fregean Senses and Russellian Propositions

The last candidates for providing access to correctness conditions for reasoning, apart from language, to be considered here are Fregean and Russellian propositions. Since Fregean propositions – also called "Fregean senses" or "thoughts$_{(F)}$" – are "[...] capable of being the common property of several thinkers" (Frege, 1960, p. 62 n), they may avoid the problems we encountered with mental representations due to their privacy.[25] But even Fregean propositions need to be grasped "[...] in an individual psychological act" (Putnam, 1975, p. 134) before they can be of any use for us in reasoning, which again leaves us only with (private) mental representations to operate on for a reasoner.

The same holds true for Russellian propositions as well. Since general Russellian propositions consist merely of propositional functions and quantifiers, their constituents share their ontological status with (the constituents of) Fregean senses as entirely abstract entities. The constituents of singular Russellian propositions, in contrast, can also be rather worldly entities, *viz.*, concrete objects. While there might be issues regarding sensitivity to intensional contexts with singular Russellian propositions, no corresponding worry exists for general Russellian propositions and Fregean propositions. However, also Russellian propositions – be they general or singular – need to be grasped by a reasoner in order to do any useful work for the context under consideration. So, the same objection we encountered regarding Fregean propositions also rules out Russellian propositions if they are to provide access to correctness conditions for reasoning: The mental representations of Russellian propositions are just as private as the representations of Fregean senses.

Even if we ignore this aspect and suppose that a reasoner can directly operate on Fregean senses or Russellian propositions without the need for a mediating mental representation thereof, Fregean and Russellian propositions are still unfit to provide access to correctness conditions for reasoning. This

[25] Fregean or Russellian propositions may be considered as the contents of mental representations, i.e., what mental representations represent. The crucial point is that mental representations are private although their content might be publicly available.

is because Fregean senses and Russellian propositions fall prey to the rule-following constraint. Unaided by language for identifying and communicating individual steps in a line of reasoning, Fregean and Russellian propositions by themselves do not allow for publicly accessible re-identification or double-checking on the legitimacy of inferential transitions between them. To achieve these features, we are cast back to language again, which is, as claimed before, the only medium fit to warrant publicly available access to correctness conditions, which is in turn necessary for reasoning.

Before we elaborate on the role language plays for reasoning (in the following section 2.2.5), and then turn to a more detailed characterization of language in general (in section 2.3), we can quickly rule out *unstructured propositions* as potential candidates to fill the role needed for reasoning (by providing access to the required correctness conditions). Unstructured propositions – e.g., propositions as functions from possible worlds to truth values or as sets of possible worlds (cf. King, 2017), in contrast to Fregean and Russellian propositions as exponents of structured propositions (cf. Rescorla, 2019, § 1.2) – are definitely too coarse-grained to do the job under consideration. Since every necessarily true claim corresponds to the same unstructured proposition (e.g., the set of all possible worlds) and every necessarily false claim also corresponds to the same unstructured proposition (e.g., the empty set), we are far from sensitivity to intensional contexts when we operate with unstructured propositions. Since the need for sensitivity to intensional contexts has already been fixed as a necessary requirement to provide access to correctness conditions for reasoning, unstructured propositions do not deserve any further attention in this regard.

2.2.5 Language in Its Constitutive Role

The point that only language can provide access to correctness conditions for reasoning and that, therefore, there is no reasoning without language can also be backed up via a slightly different route. Up to this point, I have argued that no other plausible candidate but language can provide access to correctness conditions, which are needed for reasoning. Since there is no

reasoning without the required access to correctness conditions, there is no reasoning without language. This is so far, so good, but there is an additional aspect to the story which is worth making explicit at this point.

The correctness conditions under discussion, we said, are necessary to apply (and access, as well as assess) inference rules. Since reasoning is a rule-guided activity – one that is guided by inference rules – inference rules are arguably crucial for reasoning. Yet, exactly how crucial inference rules are for reasoning can probably be best explained by making use of John Searle's (or John Rawls's, see footnote 27 below) distinction between *regulative rules* and *constitutive rules*. Searle (2013, pp. 223 f) explains the difference as follows:[26]

> I distinguish between two sorts of rules: Some regulate antecedently existing forms of behaviour; for example, the rules of etiquette regulate interpersonal relationships, but these relationships exist independently of the rules of etiquette. Some rules on the other hand do not merely regulate but create or define new forms of behaviour. The rules of football, for example, do not merely regulate the game of football, but as it were create the possibility of or define that activity. The activity of playing football is constituted by acting in accordance with these rules; football has no existence apart from these rules. I call the latter kind of rules constitutive rules and the former kind regulative rules. Regulative rules regulate a pre-existing activity, an activity whose existence is logically independent of the existence of the rules. Constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules.[27]

Applied to our topic at hand, namely reasoning, we can say that inference rules do not merely regulate how we ought to reason. Inference rules are

[26] A more or less equivalent passage appears as the first paragraph of section 2.5 in Searle (2011, pp. 33 f).

[27] Searle adds a footnote here to point to Rawls (1955) and Searle (1964), where the distinction between constitutive rules and regulative rules also appears.

therefore not merely regulative rules, but they are constitutive rules for reasoning. They create and make possible any instance of reasoning in the first place, since "[...] constitutive rules, such as those for games, provide the basis for specifications of behavior which could not be given in the absence of the rule." (Searle, 2011, p. 36) In other words, "[...] where the rule (or system of rules) is constitutive, behavior which is in accordance with the rule can receive specifications or descriptions which it could not receive if the rule or rules did not exist." (Searle, 2011, p. 35) This also holds true for reasoning, which would not be possible without (accessible) inference rules and correctness conditions.[28] "The creation of constitutive rules, as it were, creates the possibility of new forms of behavior" (Searle, 2011, p. 35).

Searle is primarily concerned with the *existence* of constitutive rules since he focuses on conventional rules which need to be brought into existence in social contexts. There is, however, no reason to assume that all constitutive rules are conventional. Inference rules and correctness conditions for reasoning are arguably not conventional but are still constitutive for reasoning.[29] In order for reasoning to come into existence, it is not sufficient that there merely *are* constitutive rules for reasoning. It is also necessary that these rules are accessible for a reasoner (in order for the individual in question to count as a reasoner; see section 3.3). Otherwise, rule-following would not be possible. We said that it is a necessary condition for reasoning that for every instance of reasoning it needs to be possible to double-check on the correctness of the line of reasoning at hand. Since the rules and correctness conditions which allow a line of reasoning to be checked for correctness are

[28] See also the quote on page 79 and the surrounding discussion (including especially footnote 26 on page 80) for elaboration on this point.

[29] Other kinds of correctness conditions certainly are conventional and even arbitrary, e.g., how to correctly move the pawn in a game of chess, how to correctly score a touchdown, or how to correctly take time out during a tennis match are arbitrary conventions, introduced at a certain point in time and potentially subject to revision. Inference rules and correctness conditions for reasoning, I hold, are neither conventional, nor arbitrary, nor subject to revision – just as the laws of logic or natural laws are not subject to revision. (Of course, our *formulations* of these laws might not be perfect yet and therefore potentially need modification when we learn something new about the laws in question which we did not know before. But that we can make scientific progress in this sense does not mean that the underlying rules and laws are conventional or arbitrary.)

only available via language, we can conclude that language is constitutive of reasoning: Not only are the rules in question constitutive for reasoning, but also access to those rules is constitutive for reasoning. As only language can provide this access, reasoning is constitutively dependent on language (possession).[30]

2.3 Language

My attempt to characterize language strongly builds on Charles F. Hockett's (1976) influential *design features* of language. Hockett's list of design features varies from seven to sixteen in the course of development of his approach (cf. Wacewicz & Żywiczyński, 2015, pp. 31 and 34). I do not accept all of Hockett's features as necessary for language. He obviously focused on speech rather than language in general. Therefore, Hockett's list includes features which exclude written language (see the third design feature in Hockett, 1976, p. 9; see also p. 14) and, crucially, even sign languages (see the first design feature in Hockett, 1976, pp. 8 f; see also Wacewicz & Żywiczyński, 2015, p. 37). Hockett's design features, in consequence, cannot successfully define the notion of language since what counts as a language according to his design features comes out too narrow (cf. Hockett, 1976, p. 15). Yet several of his features are indeed essential for a definition of language.

In order to qualify as a language, a communicative system needs to exhibit several properties. These are productivity, or a generative capacity (cf. Enfield et al., 2014, p. 6), which goes along with compositionality. Further necessary features of language are displacement (see the tenth design feature in Hockett, 1976, p. 11), syntax/grammar, and discreteness (see the ninth design feature in Hockett, 1976, pp. 10 f). Another feature which deserves attention is openness (see the eleventh design feature in Hockett, 1976, p. 11). Although openness is widely shared among paradigmatic examples of language, it should not be counted among the defining features of language. The same holds probably true for the feature of universality. Here is a list

[30] See section 3.3 for more details on the relation between reasoner, language, and reasoning.

of features which, I think, does quite a good job in characterizing what, in general, a paradigmatic language amounts to:

Discreteness of individual meaningful signs, which allows their combination to achieve syntactically more complex meaningful expressions.

Grammar which regulates how discrete signs can be combined in order to achieve the aforementioned complex expressions.

Productivity allows a language user to produce an unlimited amount of expressions from a limited amount of discrete and (more or less) primitive meaningful units together with a limited set of grammatical rules.[31] This allows a language user to potentially express an infinite amount of information.

Displacement of language allows a language user to refer to things and situations which are not present, including past, future, and fictional states or objects. This also includes things or events which are not present because they are far away, spatially speaking.

Universality, on top of displacement, allows a language user to make everything she can think of a topic of a conversation. In addition to universality, most languages are also an *open* communication system. This means that, no matter how many topics a language user can talk about, there can always be added new ones. This is an important contrast to non-human animal communication systems, which are usually *closed* and therefore not universal. (Cf. Bickerton, 1996, p. 16)

Openness is the feature of a language which makes it possible to introduce new expressions for hitherto undiscussed topics. Openness also allows

[31] Compositionality could be mentioned as the semantic side of the same coin. While productivity rather concerns the generation of signifiers, i.e., expressions, "compositionality" means the ability to generate more complex meanings, i.e., the signified, out of more primitive meanings. Decomposition of a complex and potentially novel meaning into its constituents allows the understanding – rather than the production – of infinitely many meanings by knowing only a limited amount of meanings and the finitely many rules of combination.

adaptation of "old" expressions for novel use – e.g., in metaphorical or analogical usage. This characteristic is crucial to achieve the previously mentioned universality of language.[32]

Arbitrariness of linguistic signs is a classic characteristic of language, but it may be put aside in this context in order to not exclude certain forms of language of thought [33] or alleged divine languages – e.g., the Adamic language (see also p. 100). Even if such languages do not exist, they should still count as languages if they could have existed. Such languages should therefore not be ruled out by definition.

Except for openness and (probably) universality, all of the above mentioned features are necessary for a communicative system to be a language. Taken together, the necessary features most plausibly also sum up to a sufficient condition for being a language. Most animal communication systems can be excluded from the domain of languages in this way. To the best of my knowledge, only one non-human communication system is currently known which clearly shows displacement, namely the so-called "bee language." Bees can communicate the location of food sources to their fellow bees by performing a waggle dance. So, they clearly communicate about displaced things. However, displacement in bee language has rather narrow limits. Bees can, for example, only indicate horizontal, but not vertical directions (cf. Bickerton, 1996, p. 15). And bees, as far as we know, exclusively "talk" about the location and quality of food sources. Hence, bee language is far from exhibiting the feature of universality. Still, it seems that the bee's waggle dance fulfills all necessary requirements – at least on a very basic level – to count as a language.

Should this be seen as a problematic outcome for the position defended here? Not at all, I think. The possibility of non-human languages must be acknowledged by every definition of language worthy of consideration. That only human beings apparently developed highly sophisticated communicative

[32] The feature of openness will play an important role at a later stage of the investigation in section 8.4.

[33] See BonJour's (1991) argument against the language of thought hypothesis, mentioned on page 35.

systems which allow for addressing all sorts of topics – *viz.*, which exhibit universality – is an entirely contingent matter. Is it problematic that I tend to count the bee's waggle dance among the languages even though it is arguably a rather rudimentary form of language which only just meets the necessary criteria? Does that mean that I commit myself to reasoning bees? No it does not, since I count language as a necessary, but not as a sufficient condition for reasoning. I therefore see no problem in this regard.

Inspired by the findings about the bee's waggle dance, some people tend to think that dance in general should be counted as a form of language. I disagree since I am unable to recognize any of the necessary features of language present in dance (in general) in any clear form. Leaving aside the bee's waggle dance for a moment, there are neither discrete morphemes in dance, nor is there a clear grammatical structure which builds up to productivity. I also tend to think that there is no displacement in dance in general though its absence might not be as clear as the absence of the other features.

To be clear, I do not deny that dance has expressive capacities, nor do I deny that dance can be considered as a communicative system. What I deny is that dance can be considered as a communicative system which is sufficiently elaborated and structured to count as a language. We can certainly codify dance in a way to meet the requirements for counting as a language, but then, I think, we have made a crucial step towards a sign language. Sign languages (for example ASL, i.e., American Sign Language) certainly count as full blown languages which clearly exhibit all of the listed features, not only the necessary ones. It is important to note that a sign language – being a codified communicative system with discrete meaning units and a clear syntactic structure – is different from dance. In this comparison, I would rather class dance together with so-called "body language" as clear cases of non-languages. Again, both have expressive capacities and can count as communicative systems but not as languages.

Bickerton (1996, p. 11) also mentions the fact that animal communication systems can do nothing but communicate, in contrast to human language which has additional capabilities such as storing information. But including this point among the necessary (or even paradigmatic) features

of language would come dangerously close to conflating language with an inscription system. It does not seem correct to include the presence of an inscription system in the very notion of language; otherwise merely spoken languages (which clearly are languages) might fall out of what is covered by the notion of language.

The reason why I hesitate to include universality and openness among the necessary features of language is that it might be advantageous to be able to count certain formal languages and programming languages among the languages as well. This, among other reasons, is also why I do not include pragmatic aspects among the necessary features of language. Semantic features, on the other hand, are certainly necessary for language. But this should not pose a significant obstacle for counting formal languages – e.g., predicate logic – among the languages since these languages can be readily equipped with semantic interpretations. This also gives us universality and, of course, also displacement in formal languages. (I may represent an inference about Socrates, the weather, or what ever topic I wish to address in a formal language.) The same, I think, holds true for programming languages because I cannot see any general reason why a variable or function in these languages might not represent anything whatsoever. If, on the other hand, we were to conclude that formal and programming languages should not be counted among the languages, so be it. We can simply consider these as (artificial) extensions of our (natural) languages. This would not change anything of significance for the present investigation.

The kind of languages I am primarily concerned with here – i.e., languages which can provide access to publicly available correctness conditions for reasoning – certainly exhibit all of the features listed on pp. 52 f, and more. Now that we have a more general characterization of language at hand and, by the same token, also a clearer picture of what counts and what does not count as a language, we can return to the relationship between language and reasoning. The way I set up and argued for the claim that reasoning is dependent on language will likely alarm several readers and give rise to a certain kind of concern which needs to be addressed.

Chapter 3

Further Issues

3.1 Verificationist Concerns

The discussion above, and probably especially the repeated emphasis of the rule-following constraint, might raise (anti-)verificationist concerns in some readers. This is especially reasonable because Wittgenstein's Private Language Argument was frequently accused of building on verificationist principles early on in the literature (e.g., Thomson, 1964). Since "verificationism" is a quite elastic term, it is hard to counter every accusation which might come under this label in one sweep. But the argumentation presented here is at least not guilty of a version of verificationism which clearly needs to be rejected. If the claims defended here turn on verificationist principles at all, then it is a very mild form of verificationism which can be accepted, or so I wish to argue in this section.[1]

Without giving a precise definition of what makes a verificationist position untenable or acceptable, the difference can be depicted for our purposes by considering so-called *Robinson cases*.[2] A version of the Private Language Argument with the consequence that a lonely islander like Robinson Crusoe

[1] For further and more general considerations regarding verificationism, see chapter 8 and especially section 8.3.

[2] An early account of discussing Robinson cases in connection with the Private Language Argument can be found in Ayer & Rhees (1954), where many of the mistakes pointed out below are committed.

(who never meets Friday) could not learn, speak, or understand a language because he lacks contact with a linguistic community would clearly turn on an untenable verificationist principle. The same holds *mutatis mutandis* for the Private Reasoning Argument. If if follows from *PRA* that Robinson cannot reason because no one is here to talk about his inferences with Robinson, then *PRA* would need to be rejected for unacceptable verificationist commitments. However, nothing like this follows from the arguments presented here.

Before I demonstrate this, it should be noted that the above characterization of unacceptable verificationist commitments is not a mere caricature of actually proposed verificationist positions. It would be a rather weak defense of my position to show that it does not amount to a mere straw man verificationism. It must at least be shown that *PLA* and *PRA* are also significantly different from actually proposed positions which can plausibly be deemed verificationist (and rejected on this ground). To show this, we can consider Donald Davidson's suggestion to extend Wittgenstein's Private Language Argument in a way similar to my approach. Davidson (1991, p. 157) writes:

> I believe that Wittgenstein put us on the track of the only possible answer to this question [regarding the source of the concept of truth]. The source of the concept of objective truth is interpersonal communication. Thought depends on communication. This follows at once if we suppose that language is essential to thought, and we agree with Wittgenstein that there cannot be a private language. [Footnote omitted] The central argument against private language is that unless a language is shared there is no way to distinguish between using the language correctly and using it incorrectly; only communication with another can supply an objective check. If only communication can provide a check on the correct use of words, only communication can supply a standard of objectivity in other domains, as we shall see. We have no grounds for crediting a creature with the distinction between what is thought to be the case and what is the case unless the creature has the standard provided by a shared language; and

> without this distinction there is nothing that can clearly be called thought.

This passage might, at first glance, appear very similar to what I argue for. But there are subtle, nevertheless absolutely crucial, differences between Davidson's and my account. Two apparently rather superficial differences are that Davidson talks about truth and thought (notably belief) while I talk about (inferential) correctness and reasoning. Although important, these are not the most crucial differences to be considered here. What is more important is that Davidson's formulation suggests that *actual* interpersonal communication is required for truth and thought and that only an *actually* shared language is a public and therefore possible language. If the quoted paragraph above does not merely suffer from consistently careless wording, what Davidson expresses in this passage is verificationism *par excellence*. This conception seems to make it impossible for (our version of) Robinson Crusoe to use language or even think since there is no one he could actually communicate or share a language with. I take this result to be intuitively as well as theoretically unacceptable, just as unacceptable as the apparently underlying variety of verificationism.

This kind of verificationism is, however, easily avoided. What needs to be done is to change the requirement that *actual* communication happens and that a language *is actually* shared to the requirement that communication is *possible* and that a language *can* be shared. The correct modality for those requirements is possibility, not actuality. Davidson seems to wrongly assume "[...] that unless a language *is* shared there is no way to distinguish between using the language correctly and using it incorrectly" (Davidson, 1991, p. 157; emphasis added). What Davidson should have said is that, unless a language *can be* shared, there is no way to distinguish between using the language correctly and using it incorrectly. So, a private language is not a language that *is* not shared, but a language that *cannot* be shared.

Davidson also incorrectly extends the rationale behind the Private Language Argument to the entire domain of thought, including beliefs. The consequence of this move is that, according to Davidson, a pre- or non-linguistic creature cannot have beliefs, because

> [...] to have a belief it is not enough to discriminate among aspects of the world, to behave in different ways in different circumstances [...]. Having a belief demands in addition appreciating the contrast between true belief and false, between appearance and reality, mere seeming and being. [...] Someone who has a belief about the world—or anything else—must grasp the concept of objective truth [...] (Davidson, 1991, pp. 156 f).

This in turn is only possible with language because "[...] communication is the source of objectivity [...]" (Davidson, 1991, p. 157 n) and "[...] without communication propositional thought is impossible" (Davidson, 1991, p. 160). This seems much too demanding for having beliefs. It is my conjecture that another verificationist assumption led Davidson astray here because he also claims that "[k]nowledge of our own minds and of the minds of others are mutually dependent." (Davidson, 1991, p. 160)[3] This seems to suggest – given that "[b]elief is a condition of knowledge" (Davidson, 1991, p. 156) – that the ability to individuate beliefs is necessary to have a belief in the first place.

Belief individuation might be a plausible precondition for belief *ascription*, without which there cannot be knowledge of the minds of others. But even if the requirements for ascribing beliefs to others are as demanding as Davidson claims, the requirements for merely having beliefs are much less demanding. I cannot see why a creature's inability to ascribe beliefs should block this creature from having beliefs on its own. This is probably because I do not think that a creature needs to be able to individuate (its own) beliefs in order to have them. Even if individuating beliefs is necessary for the ability to ascribe them – to other creatures and maybe even to oneself – it is implausible that merely having beliefs requires conceptual sophistication of such a high degree. Believing something does not necessarily depend on rule-following. Although we can legitimately talk about correctness conditions of beliefs, I do not need to be able (even in principle) to check on the correctness (i.e., truth) of a belief in order to have that belief. It is therefore

[3] I also think that Davidson's holistic commitments regarding belief – as expressed in Davidson (2001) – are mistaken and represent an additional defect in his account. But these considerations go beyond the scope of the present text.

illegitimate to extend the rule-following requirement to (having) beliefs and other kinds of thinking which are not reasoning. This means that I can have a belief – and what I believe might be either true or false – despite my inability to individuate the belief, discriminate it from other (maybe similar) beliefs, or re-identify said belief on later occasions.

Since I do not share any of the verificationist motives apparently present in Davidson's account and since nothing like this follows from the arguments presented and defended here, I think that verificationist concerns regarding my position are unfounded. I am, however, committed to a stance towards reasoning which might be called "restricted and mildly verificationist." Reasoning, in contrast to having beliefs and other thoughts, does require the *possibility* of an objective check of the mental transitions made. This plausibly also requires the capacity to individuate and re-identify mental contents, an ability not needed for thinking. If this amounts to a verificationist commitment, then there are (mild or moderate) versions of verificationism which can, indeed even need to be, accepted for restricted domains.[4] Reasoning is a much more demanding capacity than mere thinking, and reasoning consequentially presupposes a more sophisticated cognitive apparatus. The ability to have thoughts about thoughts, or – as Bermúdez (2003) calls it – *intentional ascent*,[5] certainly is a precondition for being able to reason. Neverthe-

[4] I will come back to this moderate and restricted form of verificationism in sections 3.3 and 8.3.

[5] According to Bermúdez (2003), intentional ascent requires semantic ascent and thereby language. Bermúdez's conclusions are therefore to a considerable extent quite similar to mine although our argumentative routes differ significantly. One crucial difference between Bermúdez's and my lines of argument is that I do not need to presuppose that metacognition or metarepresentation (i.e., having thoughts about thoughts or thinking about thoughts via intentional ascent) requires language, which is a neuralgic step in Bermúdez's argument. Bermúdez might be right in claiming that intentional ascent requires language. If so, that is fine for me although I think that there is good evidence against the hypothesis that metarepresentation requires language. Cases of purposeful deception in the animal world, for example, can hardly be explained without metarepresentation, I think – although I hold that they can be explained without reasoning in animals. (I will say more about this in the upcoming section 3.2.) This might be problematic for Bermúdez's position if we are not willing to also attribute language to an animal deceiver. So, Bermúdez's claim is certainly not trivial and might even be quite controversial. My line of argument has the advantage that I do not need to argue that language is necessary for metacognition. I might very well allow for metarepresentation without language since language as a requirement for reasoning enters my argument via

less, it needs to be clearly emphasized that by far not all mental phenomena, notably having beliefs and other kinds of thinking which are not reasoning, are so demanding.

3.2 No Non-Linguistic Reasoners?

Since reasoning constitutively depends on possession of language (as we saw in section 2.2.5), no non- or pre-linguistic being is able to reason. Yet, obviously not every kind of thought depends on language in this way. It seems fairly clear that not every kind of thought is dependent on language since non-linguistic animals and pre-linguistic infants clearly have thoughts in the broad sense sketched in section 1.3. Also experiences are language-independent and can therefore be had by all sorts of conscious beings even if they clearly lack language.[6] It may also be worth pointing out at this point that a lack of language does not imply a complete lack of communicative abilities. It should go without saying that a plethora of species communicate without possessing anything that comes even close to language.

Also human communication is often not entirely linguistic, as can be seen from, e.g., so-called body "language." I add scare quotes here because I am not prepared to accept body language as a language in the literal sense (cf. Enfield et al., 2014, p. 6). Body language is without any doubt a means of communication, but what is communicated in this regard cannot be considered as linguistically encoded information, except for cases where a message is conveyed on the basis of an established code. Then, however, we are not speaking about body language anymore but instead about *sign languages* (e.g., ASL) which are proper, full-blown languages (cf. Ahearn, 2017, pp. 45 f). It can also be added at this point that "[. . .] *body language* as used by humans suffers the same limitations as *'animal languages.'*" (Bickerton, 1996, p. 12) So, the reason why body language is not a language partially

an entirely different route – namely by way of the ability to check a line of reasoning for correctness, not in order to have metacognition *per se*. See p. 75 for further details.

[6] For more on the relation between language and (the expression of) experience, see section 9.2.

overlaps with the reason why so-called "animal languages" are not languages.[7] "Because so many people confuse language with communication, [...]" (Bickerton, 1996, p. 11) it is important to keep these truisms in mind when we try to evaluate how plausible or convincing the position presented here might be. However, a clear consequence of *PRA* is that non-linguistic individuals – although they can think, believe, feel, and have experiences – cannot possibly be said to reason insofar as they lack a constitutive element of reasoning. In other words, there are no pre- or non-linguistic reasoners.

Still, it needs to be emphasized that this does not necessarily mean that there is no non-linguistic reasoning. This caveat is neatly brought to light by Martine Nida-Rümelin (2010), who clearly distinguishes two questions:

> (Q1) Is it possible for a creature without a language to [reason]? (Can non-linguistic creatures [reason]?)
>
> (Q2) Is it possible to [reason] without thereby using a language? (Nida-Rümelin, 2010, pp. 55 f)[8]

While it clearly follows from *PRA* that (Q1) needs to be answered in the negative, the same answer is not necessarily implied for (Q2). (Q2) should be read as asking whether linguistic creatures can reason without employing language.[9] This possibility is not ruled out by *PRA*, given the interpretation argued for in section 3.1. The moderate and restricted form of verificationism adopted there merely demands that it be *possible* to communicate a line of thought for it to qualify as reasoning. This is compatible with a non-linguistically framed line of reasoning as long as it can be put into language in order to be communicated and (re-)evaluated. In this way we have a clear sense in which non-linguistic creatures cannot reason although there is non-linguistic reasoning since even non-linguistic reasoning is unavailable to non-linguistic creatures.

[7] See also section 2.3, where these issues are discussed in some more detail.

[8] The expression "think" was changed to "reason" in this quote.

[9] An example for this kind of reasoning – namely imagistic reasoning – was already considered in section 2.2.3.2.

The possibility of non-linguistic reasoning might suggest to some readers a commitment to non-conceptual content. While I am aware that it might be tempting to read the previous paragraphs as inviting a position about reasoning which is sympathetic to non-conceptual reasoning, I would warn against reading any non-conceptual commitments into the position defended here. I tend to be rather skeptical regarding non-conceptual mental content in general since I follow a low-level approach to concept possession. This means that being able to sort, e.g., bricks according to their color is sufficient for having the concept COLOR.

Many non-conceptualists, in contrast, prefer a high-level theory of concept possession, according to which this ability is not sufficient for concept possession. This complicates a comparison of non-conceptualist theories with the account defended here. Be that as it may, I wish to emphasize that no commitment regarding non-conceptual content, neither positive nor negative, directly follows from the view presented here.[10] Still, I rather think that even non-linguistic reasoning probably cannot be non-conceptual, and, even if non-linguistic reasoning might not be fully propositional, I suppose that it still needs to be conceptual. The notion of non-conceptual content is therefore irrelevant for the present discussion and can consequentially be ignored in this context. Let us instead return to non-linguistic creatures.

[10] I hope to elaborate on the reasons for my skepticism regarding (the usefulness of the notion of) non-conceptual content elsewhere in the future, but I will not go into details concerning this matter now. I will just state how I see the situation regarding the position defended here in very rough terms: I think that the distinction between 'the state view' and 'the content view' of non-conceptual content (cf. Heck, 2000, p. 485) is crucial. If non-conceptual content is understood along the lines of the *state view,* there is no conflict whatsoever between the position defended here and non-conceptual content. If non-conceptual content is understood along the lines of the *content view,* however, then my position might be incompatible with non-conceptual content. (Some details which exceed the scope of the present investigation need to be settled before we can conclusively say whether non-conceptual content in this sense is really in conflict with the position defended here.) All the same, I think that non-conceptual content in the sense of the content view *on the conscious level* is illusory. Non-conceptual content construed along the content view must, on my view, be consciously inaccessible. Although I do not have the space to provide arguments for this position here, the incompatibility of my position with the assumption that there is conscious non-conceptual content in the content view sense is, I think, unproblematic since it is just an incompatibility with a wrongheaded and therefore mistaken account of content.

If it is true that non-linguistic creatures are not able to reason, how are we to account for the partly impressive cognitive achievements and problem-solving abilities in the (non-human) animal world? Some clarifications are in order before this question can be seriously considered. First of all, the claim that there are no pre- or non-linguistic reasoners is not equivalent with the claim that there are no non-human reasoners. At least conceptually, the domains of humans and reasoners do not coincide insofar as neither notion – i.e., *being human* and *being able to reason* – is implied by the other. To be a human being is not sufficient for being a reasoner since pre-linguistic infants are not among the reasoners although they clearly belong to the human species. Moreover, speaking non-human creatures could perfectly well be capable of reasoning, so being human is also not necessary for reasoning.

The most prominent case of (alleged) animal reasoning in the history of philosophy is probably the story about *Chrysippus' dog*: A dog, hunting its prey, comes to a tripartite crossroad. The prey must have taken one of the three pathways. The dog sniffs at the first path but does not pick up the prey's scent. So the dog sniffs at the second road and again does not smell the quarry. Therefore, the hunting dog rushes down the third path without even sniffing for the prey in this direction. This scenario is supposed to suggest that the dog reasoned as follows: The prey took one of the three roads, it did not take the first, and it did not take the second, thus it took the third. Reasoning along these lines constitutes full-blown reasoning indeed since it is an instance of *following* a disjunctive syllogism. But according to the position defended here, a dog (given that it does not have language) could not engage in reasoning at all. So it could not possibly reason through a disjunctive syllogism.

The story of Chrysippus' dog is an ancient philosophical problem case and has provoked various reactions. Some philosophers have been happy to ascribe reasoning to animals while others have tried to defuse the story of Chrysippus' dog in several ways without conceding deductive capacities to animals. I will neither discuss traditional reactions to Chrysippus' dog, nor will I develop a new account. I just wish to discuss Chrysippus' dog as an exemplary case to show how apparent problem cases can be accommodated

to the position defended here. To this end, we can cite Rescorla (2009) who presents a solution to Chrysippus' dog which allows an explanation of the dog's behavior (and cognitive achievement) without attributing deductive reasoning to the animal. This is done, in short, by constructing a Bayesian probabilistic decision model, operating on mental maps. Leaving the intricate technical details aside, we can conclude with Rescorla (2009, p. 67) that "[...] a satisfying treatment of Chrysippus' dog need not cite logical reasoning over logically structured mental states [...]" because "[t]he relevant processes, grounded in Bayesian decision theory, differ markedly from deduction." (Rescorla, 2009, p. 71) This means that we can provide a model that predicts the dog's behavior without attributing cognitive capacities to the dog it cannot have, according to my account, without possessing language. Since language and reasoning are not necessary to explain the dog's behavior, Chrysippus' dog is defused as a counterargument against *PRA*. Note, however, that this line of argument builds on "[...] a basic methodological tenet of the science of animal behavior called Lloyd Morgan's canon:" (Lurz, 2009, p. 7)

> *In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale.* (Morgan, 1896, p. 53)

Morgan's canon basically represents a specialized version of Occam's razor for the domain of comparative psychology (cf. Andrews & Monsó, 2021, § 2.3), but it is not universally accepted – or at least its correct interpretation is often hotly debated. However, if we can legitimately make use of Morgan's canon in this case, then having the cognitive abilities to act according to (without *following*) a Bayesian model is sufficient to explain Chrysippus' dog but does not require that Chrysippus' dog is able to reason, e.g., by drawing a disjunctive syllogism and thereby following the corresponding rules.[11]

Rescorla's (2009, p. 58) "[...] Bayesian-cum-cartographic model of Chrysip-

[11] I presuppose that the capacity to follow a rule stands higher on the psychological scale than the ability to merely act according to a rule.

pus' dog" can explain what happens in the story and "[...] countenances non-linguistic cognition while sharply distinguishing it from linguistic cognition." (Rescorla, 2009, p. 53) In other words, Rescorla (2009) satisfactorily explains Chrysippus' dog without positing that the animal is able to draw deductive inferences, which is a kind of reasoning and would therefore require language possession. Although the cognitive model to explain Chrysippus' dog might be quite complex, it merely requires cognitive processes which are *in agreement* with (e.g., Bayesian probabilistic) rules, in contrast to cognitive processes which constitute rule-following.[12] Complex cognitive models are needed in any case to explain most mental processes which are language-independent, e.g., visual processing. What is crucial here is that rule-following is not needed to exemplify these models. Therefore, these non-linguistic cognitive processes can be executed unconsciously (i.e., on the sub-personal level) and do not contradict the claim that *conscious* reasoning – which is the only kind of reasoning I admit – is language-dependent.[13]

3.3 What Makes a Reasoner? – Outline of an Ability-Based Approach

In order to qualify as a reasoner, a subject needs to be able – at least in principle – to check on the correctness of her line of reasoning. I take this to be a conceptual truth about reasoning: Every line of reasoning is either correct or incorrect, and in order to qualify as a line of reasoning, it needs to be possible – at least in principle – to figure out whether a given line of reasoning is correct or not. A line of reasoning which cannot, even in principle, be qualified as correct or incorrect or to which the notion of correctness/incorrectness does not even apply – for whatever reason –[14] is impossible.

[12] I will come back to Chrysippus' dog on pp. 76 ff.

[13] The topic of animal cognition is a vast and intriguing field which cannot be further treated at this point. Many pressing questions therefore need to remain open in this context. I therefore hope for an occasion to address challenges coming from studies of animal cognition for the account presented here in a less cursory form in the near future.

[14] Remember, for example, word association tasks (mentioned already on page 16), where it does not make sense to qualify the underlying cognitive thought processes as

This requirement relates back to the moderate and restricted verificationism I endorse, introduced in section 3.1. This version of verificationism is *moderate* because it merely demands that it is possible to settle the question of correctness, not that the question needs to be actually settled or even that it needs to be clear how to settle whether a line of reasoning is correct, for example by having a method of verification at hand. And this version of verificationism is *restricted* because it is limited to the domain of reasoning, and does not apply to the entire domain of thought.

Still, in order to answer the questions of what qualifies as a line of reasoning and who qualifies as a reasoner, more needs to be said about what it means that it is *principally* possible for a subject to check on the correctness of a line of reasoning. The expression "principally" or "in principle" introduces a substantial amount of indeterminacy which needs to be resolved. In order to answer some pressing questions in this context, let us first take a step back and restate some basic facts to clear the ground.

We are concerned with (lines of) thought and the question of what makes (a line of) thought be (a line of) reasoning instead of merely thinking. In order for (a line of) thought to qualify as reasoning, it needs to be possible that the (line of) thought in question can be evaluated for a certain kind of correctness. So the question "What makes reasoning?" amounts to the question "What does it take to make something a candidate for evaluation of correctness in the relevant sense?", where the relevant or right sense of correctness is akin to validity but not truth or other possible kinds of evaluation which might also legitimately be called "correctness."

So far, this is only a repetition of what has already been said before. In order to avoid repeated wordy clarifications regarding which type of correctness is at issue, I will use the expressions "(in-)correctness" or "(in-)correct" with the subscript *"i"* for inferential correctness akin to validity and "(in-)correctness" or "(in-)correct" with the subscript *"t"* for truth-conditional correctness. So, being evaluable for correctness$_t$ – as, e.g., beliefs are – does not make a thought an instance of reasoning. In order to count as reasoning, a thought

either correct or incorrect although other norms of evaluation might apply, e.g., normal or pathological.

needs to be evaluable for correctness$_i$. But which conditions need to be fulfilled in detail to be either correct$_i$ or incorrect$_i$?

A first aspect we need to settle is which extent a thought needs to have, in order to qualify as reasoning. An isolated thought – e.g., a belief, such as the belief *that Napoleon won the Battle of Waterloo* – does not qualify for correctness$_i$. If we take this thought in isolation, it does not make sense to ask whether it is inferentially correct, i.e., whether the thought *that Napoleon won the Battle of Waterloo* is correct$_i$ or incorrect$_i$. Only a line of thought which culminates in this thought – or the negation of this thought or any other isolated thought – could possibly be (in-)correct$_i$. The thought in isolation qualifies as (in-)correct$_t$ but not as (in-)correct$_i$. In fact, the thought that Napoleon won the Battle of Waterloo is incorrect$_t$ since Napoleon was defeated at Waterloo. The crucial point is, however, that, in order for a line of thought to qualify as (in-)correct$_i$, it needs to be embedded in an appropriate context. This can be stated bluntly: In order to be inferentially correct or incorrect, i.e., (in-)correct$_i$, a line of thought needs to qualify as an inference. If it does not qualify as an inference, it cannot be evaluated for correctness$_i$.

The point emphasized here is perfectly analogous to the fact that validity does not apply to sentences and truth does not apply to arguments. Any attempt to evaluate an entire argument for truth instead of evaluating its premises or its conclusion in this way, as well as any attempt to evaluate an individual sentence for validity (in the sense of being deductively, inductively, or abductively correct), is simply a category mistake.[15] Although I take deductive validity to be the paradigmatic example for validity and most of the present considerations are formulated with deductive validity in mind, I do of course recognize that deductive validity is not the only kind of validity. Correctness$_i$ is meant to also comprise cases of inductive (in-)validity, as well as abductive (in-)validity, and potentially also other kinds of (in-)validity,

[15] I ignore the fact that logical truths – as in the case of a thought which has a content of, e.g., the logical form $(P \vee \neg P)$ – are sometimes called "valid." I reserve the expressions "valid" and "validity" (as well as "invalid" and "invalidity") exclusively for *inferential* correctness. Therefore, individual sentences or the contents of individual thoughts cannot be valid in this sense.

such as probabilistic (in-)validity, for example.[16]

I take a pluralist stance regarding the notion of correctness$_i$, which means that the question "Is line of reasoning X correct$_i$: Yes or No?" is harder to answer than it might initially seem. For many lines of reasoning, "yes *and* no" seems to be the correct answer. A given line of reasoning might, e.g., be deductively invalid but inductively valid. Since correctness$_i$ is meant to apply to all kinds of reasoning and since it is a fundamental aspect of (every kind of) reasoning that it can be evaluated for correctness$_i$, we should find an account which allows a determinate answer to the question regarding correctness$_i$ for every instance of reasoning.

To achieve this, it is necessary to admit that inferential correctness *tout court* probably does not exist. Any question regarding the correctness$_i$ of a given line of reasoning can only be answered against the backdrop of a fixed standard of evaluation. Are we, e.g., to evaluate a line of reasoning for deductive correctness or inductive correctness? Without fixing the standard of correctness – which is different for deductive and inductive inferences – no determinate answer to the question "Is this line of reasoning inferentially correct?" can be given.

What makes the situation even worse is that, while correctness$_i$ against the backdrop of deductive validity might plausibly allow a yes-or-no answer, the situation is not so clear regarding other kinds of inferential correctness. Inductive, probabilistic, and especially abductive inferences probably cannot be qualified as correct$_i$ or incorrect$_i$ without stipulating a more or less arbitrary boundary between validity and invalidity. An evaluation of abductive inferences, for example, plausibly does not permit a clear-cut distinction between correct and incorrect inferences. Rather, it might be the case that the best we can do is to sort abductive inferences into better and worse specimens, *viz.*, stronger, more credible, and better confirmed vs. weaker, more

[16] I wish to stay neutral regarding the question how many fundamental kinds of inference there are. I tentatively follow the standard view that there are at least three fundamental kinds: deduction, induction, and abduction. Probabilistic inference may or may not be a sub-type of inductive inference, and the same holds true for abductive inference. So, the question of whether there are two, three, four, or even more irreducible kinds of inference – and a corresponding number of irreducible kinds of validity – is irrelevant for the point in question.

brittle, and shaky inferences. So, abductive (and potentially also other kinds of) validity is a matter or more or less, not of yes or no.

I am therefore prepared to admit that a unified account of validity – and, by the same token, of (in-)correctness$_i$ – probably cannot be found. However, pluralism regarding inferential correctness does not imply logical pluralism, i.e., the view that there is not only more than one correct logic in the sense that a framework for evaluating inductive inferences must be different from a framework for evaluation of deductive inferences, but that there are competing logical frameworks (for the same domain of application) which can be equally correct although they give conflicting answers to the question of whether a certain line of reasoning is correct$_i$ or not. Let me give one example for illustration: Leibniz's modal ontological argument (for the existence of God) can be qualified as valid in S5, but it is invalid in S4 (cf. Look, 2018, p. 707). While I am prepared to admit that we cannot do better than saying that a given line of reasoning is, e.g., deductively incorrect but nevertheless inductively correct, I think there must be a more conclusive answer to be found to the question of whether Leibniz's modal ontological argument is correct than merely saying that it is correct according to S5 but incorrect according to S4. There must be a definitive answer to the question whether Leibniz's modal ontological argument is (deductively) correct or not, full stop.[17]

However, these questions cannot be further pursued here since subsequent questions such as "Is there *the* true logic, and if so, which one is it?" and "Are there 'valid(ity)-makers' (in analogy to truth-makers), and what are they?" exceed the scope of the present investigation by far.[18] I therefore can, in this

[17] A crucial part of an answer to this question needs to consist in an answer to the question of whether metaphysical modality does indeed collapse, as expressed in the decisive axiom of S5 ($\Diamond \Box \varphi \rightarrow \Box \varphi$) (cf. Look, 2018, p. 707). If metaphysical modality does behave according to this axiom, then S5 is indeed the correct logical framework to evaluate arguments concerned with metaphysical modality, and consequentially we need to come to the conclusion that Leibniz's modal ontological argument is valid, full stop. If, however, metaphysical modality does not collapse as indicated in the axiom in question, then a weaker system than S5 turns out to be correct, and we know that Leibniz's argument is in fact invalid.

[18] I offer one word about the question regarding valid(ity)-makers for arguments or lines of reasoning (in analogy to truth-makers for sentences or beliefs): A common answer

context, only wave towards my strong realist inclinations regarding logic, which can be upheld in the face of a pluralist stance towards the notions of validity and correctness$_i$.[19]

Let us come back to considerations which lie closer to the issues at hand and which (in contrast to the aforementioned perennially problematic questions in philosophy of logic) need to be definitely answered in the context of the present investigation. Lines of reasoning – or reasonings as I will sometimes call them for the sake of brevity – need to consist of at least two individual thoughts just as an argument needs to consist of at least one premise and a conclusion. A further question is how elaborate and explicit a line of thought needs to be in order to count as reasoning. The individual thoughts which serve as premises of a reasoning need not be explicitly represented *as* premises by a reasoner. The same holds true for the result or endpoint of a line of reasoning, i.e., the thought a reasoning culminates in: in other words, the conclusion. The result of a reasoning does not need to be explicitly represented *as* a conclusion by the reasoner. It is likewise

which seems to be readily at hand is to say that what makes an argument (or a line of reasoning) valid is that it conforms with (the inference rules of) a given system of logic, e.g., Classical logic, Paraconsistent logics, or different systems of modal logic, such as S4 and S5. But we can construct all sorts of logical systems, with all sorts of inference rules, which might give utterly absurd answers to the question of which kinds of inferences are valid and which are not. So, merely pointing to some logical system(s) and saying that they make inferences valid or invalid must be entirely unsatisfactory as long as we do not have an independent account which tells us why some systems are "better" than others. (By "independent" I mean here independent of our mere logical intuitions, which tell us that some inferences are plainly incorrect$_i$ and that a system which has it that such an inference is valid must therefore be wrong.) Also, my realist leanings commit me to the view that, e.g., Leibniz's modal ontological argument must already have been correct or incorrect (whichever it actually is) before systems like S4 and S5 got invented. A mere instrumentalist stance towards logical systems, I think, cannot be the last word in this debate. Either logical systems (in some way) "track" logical facts, or they do not. If they do not track independent logical facts, they cannot be "the true logic," and if they do, then it is these logical facts (which exist independently of all the logical systems we came up with) which make an argument or line of reasoning valid or invalid, not the systems themselves. Unfortunately, I do not have a clear idea at the moment what these logical facts might be and which ontological status they have. These questions therefore strike me as equally pressing and puzzling. I will, however, leave things at that for the moment since this is not the place to dive into the metaphysical depths of philosophy of logic.

[19] I wish to thank Thomas Raleigh for illuminating discussions about these topics. While, I think, he does not fully share my outlook on these issues, he has definitely helped me to see the problems much more clearly.

not necessary that a reasoner explicitly represents the inference rule(s) which lead from one step of her reasoning to the next.[20] A reasoner does not even need to be aware that she followed rules in drawing her inference; much less does she need to represent exactly which rule(s) she followed or what makes her application of a rule correct or incorrect.

Raising such requirements for reasoning would amount to an extreme overintellectualization of reasoning. Awareness of how reasoning proceeds and of what constitutes reasoning is not required to engage in reasoning. The vast majority of reasoners might have never thought about inference rules, let alone that a reasoner would need to be able to cite individual inference rules like *modus (ponendo/tolendo) ponens* or *modus (tollendo/ponendo) tollens*. Obviously, nothing of that can be a requirement for reasoning since otherwise we would need to claim that only people who have attended a logic course, or a similar explicit training, can count as reasoners. This would be an absurd claim to make. Graduating from an explicit training in logic or critical thinking can certainly improve reasoning, but it cannot be a prerequisite for reasoning.

Neither actual awareness of which rules were applied in a certain line of reasoning nor the generic awareness that rules were applied at all is necessary for reasoning. An intuition regarding correctness or incorrectness of an inference drawn or a rule applied is also not necessary for reasoning. These kinds of metacognition do not need to accompany an instance of reasoning. A psychologically modest account of what it takes for a line of thought to be reasoning, and of which accompanying metathoughts are necessary, is not only due to theoretical reasons. The requirement of representing applications of inference rules in addition to the mere application of rules (without

[20] I assume that it is possible to follow a rule without explicitly representing the rule being followed (or the following of the rule), just as we follow grammatical rules while speaking without explicitly representing the rules we follow. Most speakers who are perfectly able to follow grammatical rules are not in a position to explicitly represent the rules they follow. Even when confronted with possible formulations of the rules they follow, most speakers would perform rather poorly in identifying the rules which actually govern their linguistic behavior. But this is unproblematic since "[...] speakers do not have to know explicitly the rules they follow in order to follow them." (Navarro-Reyes, 2009, p. 297; drawing on Searle, 1995, pp. 139 ff). See also Searle (2011, pp. 41 f) for a similar illustration of the point in question.

explicit representation of the fact that they were applied) might lead into a vicious regress, as Lewis Carroll (1895) famously demonstrated. Just as a rule of inference does not need to appear as a premise in an argument, an inference rule also does not need to be represented by a reasoner even if the reasoner in question actually applies the rule of inference in question. Apart from this theoretical aspect, we would do better to look for a psychologically realistic account of reasoning since it seems simply implausible that reasoning always requires additional representations of the functioning of reasoning among the reasoner's occurrent thoughts. As I have said, disregarding the looming regress, such an account of what it takes to reason is also psychologically incredible. Modesty is therefore also called for regarding which kinds of metarepresentation are required for reasoning.

Many creatures which are not able to reason are arguably also unable to represent metacognitive facts. They cannot, for example, represent the fact that they are drawing an inference, nor can they represent aspects of how they draw an inference. But actual presence of such additional metarepresentations is not necessary for reasoning in any case. In order for a conscious mental activity to count as rule-governed, it is not necessary that its rule-governedness also needs to be consciously represented. I grant, of course, that "[. . .] there is a difference between inferring a conclusion which follows from your premises and inferring a conclusion *because* it follows from your premises." (McHugh & Way, 2015, p. 137) I also grant that "there is a difference between conforming to a rule and doing so by being sensitive to it." (McHugh & Way, 2015, p. 139, n. 8) In order to reason, it is certainly necessary that the reasoner not merely conforms to a rule but also that she follows it. In order to follow a rule, it is definitely necessary that the reasoner be sensitive to the rule. But being sensitive to a rule does not include an explicit or conscious representation of the rule being followed nor a conscious mental representation of the fact that the reasoner is following a rule. Still, in order to count as being sensitive to a rule, a reasoner must at least be capable of representing the rule even if the rule is never actually represented.

However, although an actual representation of such metacognitive facts is not necessary for reasoning, the *capacity* to represent them is required.

Without the ability to (meta-)represent one's own mental contents and the possible transitions between them, the correctness of an inference cannot be evaluated. Although a capacity for metarepresentation in this sense is certainly necessary for reasoning,[21] it is not sufficient. A reasoner also needs to be able to represent metacognitive facts in a certain way, namely in a way which allows intersubjectively available evaluation of correctness$_i$. In order to make clear what it means to say that it is in principle possible for a subject to check on the correctness of a line of reasoning, we can take recourse to three *core abilities* which are necessary (and together, I think, sufficient) to count as a reasoner:

1. language possession,[22]

2. general metarepresentational capacities, and

3. a general understanding of the fact that inferences can be correct or incorrect (or at least better or worse).[23]

These core abilities are a weaker requirement than the *general ability* – usually associated with skills and know-how (cf. Hofmann, 2021, p. 10) – to check reasoning for correctness since having the ability to metarepresent all kinds of facts about a line of reasoning does not mean that these facts actually need to be represented while reasoning. It also does not mean that these metarepresentations need to be readily at hand for a certain reasoner. It might take a considerable and even indeterminate amount of reflection for a reasoner to identify the crucial aspects of a given line of reasoning which make the line of reasoning correct$_i$ or incorrect$_i$. The core abilities which constitute a reasoner are also weaker than a general ability to check on the correctness of reasoning because they do not include knowledge how to *actually* check a line of reasoning for correctness. Yet, with the core abilities at

[21] See footnote 5 on pp. 61 f.

[22] For the reasons discussed in section 2.2.

[23] So I think an analogous fact of what Davidson claims regarding belief – that having a belief requires "[...] appreciating the contrast between true belief and false [...]" (Davidson, 1991, p. 156); see also p. 60 for the quote in context – does actually hold true for reasoning, but not for believing. See also footnote 5 on page 26.

hand, an individual can in principle figure these aspects out – even without explicit training in (formal) logic – and can therefore count as a reasoner.[24]

It should also be noted that a subject is not necessarily in a privileged position to check on the correctness of her own lines of reasoning. A given reasoner might lack the knowledge, attentiveness, or willingness to check on the correctness of his reasoning. Somebody else might be in a much better position to actually determine whether a given line of reasoning was correct or not. But as long as a reasoner has the abovementioned core abilities, she can count as a reasoner because she can at least in principle evaluate a line of reasoning for correctness. And, crucially, having the core abilities listed before is also necessary for a reasoner to be sensitive to a rule (of inference). So, having these core abilities also makes it possible for a subject to actually *follow a rule*, instead of merely conforming to – i.e., acting (or thinking) in accordance with – it.

Non-linguistic creatures are excluded from the domain of reasoners not only due to their restricted metarepresentational capacities. Hardly any metarepresentation needs to be actually employed, on my view, during reasoning. It could also happen that a non-linguistic creature and a reasoner have the very same line of thought in mind and that this line of thought counts as a line of reasoning in one creature while it cannot be a line of reasoning in the other, i.e., the non-linguistic, creature. This result might seem counterintuitive. How can the very same line of thought be reasoning in one case but not in the other? This question deserves some elaboration.

In order to make the discussion a little less abstract, let us take Chrysippus' dog as an example again. Remember that Chrysippus' dog comes to a tripartite fork along its way while hunting. Let us assume the following statements capture what passes through the dog's stream of consciousness in

[24] Being in the know about which aspects of a line of reasoning make it correct$_i$ or incorrect$_i$, and knowing which are the right inference rules to use in certain situations – i.e., the kind of knowledge one acquires by going through a training in logic – would amount to having the *narrow ability* (cf. Hofmann, 2021, p. 10) to represent all kinds of facts about a given line of reasoning and potentially also all sorts of metarepresentations about one's reasoning. But the demand that one needs to have such narrow abilities would amount to an untenable overintellectualization of reasoning. The much less demanding core abilities are, as I said, sufficient to count as a reasoner.

sufficient accuracy:

1. The prey took one of the three paths.
2. The prey did not take the first path.
3. The prey did not take the second path.
4. The prey took the third path.

Chrysippus' dog thinks (1) when it comes to the crossroad, it thinks (2) after sniffing in one possible direction the prey might have taken, and it thinks (3) after sniffing in another possible direction. Finally, a content amounting to (4) appears in the dog's mind, and the animal continues its hunt by following the third path. According to the traditional story, the dog runs along the third path without sniffing in this direction for confirmation. For the purpose at hand, however, this detail is irrelevant.

We already discussed how the dog's behavior and cognitive achievement can be explained without attributing reasoning to the dog in the previous section on pp. 65-67. The question here is: How can it be that this line of thought counts as reasoning when it passes through my mind but merely as thinking in the dog's mind? The assumption is that it is the exact same line of thought which passes through my and the dog's minds. In order to make this assumption plausible, we cannot presume that these thoughts pass through my stream of consciousness in verbalized form as if I told myself (1) through (4) silently in inner speech. The dog arguably cannot represent (1) through (4) in this way. The dog's and my mental contents, *ex hypothesi*, need to match. The mental content in question therefore cannot be linguistically represented. Since I explicitly leave the possibility of non-linguistic reasoning open (see pp. 63 f, as well as section 2.2.3.2), this restriction is unproblematic in this context. The line of thought under consideration here needs to present the information in (1) through (4) in a way which allows that the dog and I have the same in mind. I assume, for the sake of argument, that a format exists in which the relevant information can be presented for the dog and for me in the same way.

To achieve the required sameness of content in the dog's and my mind, we also need to assume that the relevant information is presented exclusively in concepts which the dog and I share. (1) through (4) can only be very rough approximations to these requirements. I have already mentioned that the relevant content must not be linguistically framed since I presuppose that a dog is a non-linguistic creature which cannot represent information in a linguistic form. Also my notions of PREY and PATH might be too rich for the dog to have them. The same problem might appear with (ordinal) number concepts – such as THREE and SECOND – which play a role in how I presented (1) through (4) above. Also in these regards I assume that a common denominator can be found which lets the dog and me have the very same line of thought in mind. With all the caveats mentioned already, (1) through (4) can only be an extremely rough and inaccurate approximation to what this content, which the dog and I share, actually is.

Be that as it may. Given that we can find a "downgraded" version of the line of thought in question, which a dog can just as easily entertain as I can, I still claim that this line of thought will be reasoning in my mind while it merely qualifies as thinking in the mind of the dog. Since we assume that the dog's and my mental content is identical while we share the relevant line of thought, it cannot depend on the content of the line of thought in question whether it qualifies as reasoning or not. In fact, I think that being (a line of) reasoning is not an intrinsic property at all. To say that being (a line of) reasoning is not an intrinsic property means that whether a line of thought qualifies as reasoning does not depend on the line of thought in question. Being (a line of) reasoning, I maintain, is a relational property. Whether something is (a line of) reasoning depends on factors "external" to the line of thought. Whether a line of thought is reasoning or thinking depends on aspects of the mind which holds the line of thought, not on aspects of the line of thought itself. This also commits us to the view that the property of being (a line of) thinking and the property of being (a line of) reasoning cannot be attributed to sequences of mental contents conceived of as types of lines of thought. Only *tokenings* of lines of thought can be characterized as either instances of thinking or reasoning, depending on the capacities of

the mind where the line of thought is instantiated.

Why should we believe that the property of being reasoning is a relational and not an intrinsic property? I claim that the crucial characteristic of reasoning is that it can be evaluated for correctness. A line of thought needs to fulfill certain requirements to be evaluable for correctness. It needs to be of the right kind to be inferentially correct, i.e., correct$_i$, not correct$_t$. This means especially that a thought, in order to count as reasoning, needs to be *a line of* thought, not an isolated belief, for example. The latter could only be correct$_t$ whereas only the former could be correct$_i$. It is a necessary requirement for reasoning that it be evaluable for correctness$_i$. This means – as repeatedly emphasized – that for every line of reasoning it is at least in principle possible to figure out whether it is correct$_i$ or not. That it is at least in principle possible to evaluate a reasoning for correctness means that the reasoning appears in a mind with the capacities required for evaluating a reasoning for its correctness. And a crucial capacity a minds needs to have in order to be able to evaluate a reasoning for correctness is language (together with the other core abilities) since only language provides access to the intersubjectively available correctness conditions against which a line of reasoning can be checked for correctness.[25]

The situation regarding the dog and me, having the same line of thought in mind, is similar to a situation Searle (2011, pp. 35 f) brings up:

> It is possible that twenty-two men might go through the same physical movements as are gone through by two teams at a football game, but if there were no rules of football, that is, no antecedently existing game of football, there is no sense in which their behavior could be described as playing football.

We have an analogous situation with the dog and me: Although the dog and I have the same line of thought in mind – *viz.*, we go through the same mental

[25] We might object at this point that it is sufficient if the reasoning in question can be checked for correctness by anyone, not necessarily by the subject who entertains the line of reasoning. But if a reasoner has the abilities to make a line of reasoning intersubjectively available, so that it can be checked for correctness by someone else, then the reasoner also has the abilities to check on the correctness of her line of reasoning on herself, at least in principle.

"movements" – it is still not the case that the dog reasons whereas I reason. Just as the twenty-two men do not play football though the corresponding two teams do play football. In this example, Searle apparently imagines a scenario where there is no game of football – and correspondingly no rules of football – presumably because it was never invented. This is a disanalogy between our case at hand and Searle's football example. The reason why the dog is not reasoning although I am reasoning is obviously not that there are no rules of inference or correctness conditions.[26] The reason is that I have access to correctness conditions for reasoning while the dog does not have this kind of access. It is not only necessary that inference rules and correctness conditions exist in order for there to be reasoning. It must also be possible for a reasoner to have access to the correctness conditions in question. Otherwise, there is no reasoning going on. It is due to certain abilities I have, but which the dog lacks, that I am a reasoner whereas the dog is a non-reasoner because the dog is a non-linguistic creature.

Let us recapitulate: It is not the actual transitions a reasoner goes through which make her a reasoner, and the line of thought into a line of reasoning, but the capacity to represent and reconsider the transitions made. It is therefore possible that a linguistic creature and a non-linguistic creature happen to have exactly the same line of thought in mind, and, while this line of thought counts as reasoning in the mind of the linguistic creature, it does not count as reasoning in the mind of the non-linguistic creature. This consequence can be put as a slogan: *It is not the reasoning which makes the reasoner, but it is the reasoner who makes the reasoning.* This slogan can be read in several different ways, but in this context it means the following: We cannot qualify a sequence of mental contents as reasoning or not reasoning *per se*. Whether a line of thought is reasoning or not depends on the capacities of the individual which happens to have this line of thought. So, whether a creature is a reasoner cannot be determined by looking at the lines of thought this creature has. It is not the thoughts a creature has which make it a reasoner.

[26] It is not even clear whether a counterfactual situation, analogous to the one apparently imagined by Searle, is even possible for the case of reasoning. Inference rules and correctness conditions might exist necessarily, in contrast to the rules of football, which obviously exist contingently.

It is the creature (and the capacities it has or lacks) which makes a certain line of thought an instance of reasoning. In other words, you first need to be a reasoner in order to reason. It is not the case that reasoning makes you a reasoner. Reasoning is ontologically dependent on a reasoner, not the other way round.

The way I set up this discussion might suggest that I am committed to a position which Matthew Boyle (2016) calls an *additive theory*, in contrast to a *transformative theory*.[27] Boyle (2016) is concerned with a notion I, for the most part, avoided in this investigation, namely rationality. Boyle introduces the distinction between additive and transformative views as a distinction among theories of rationality. So his focus is certainly different from mine. But the distinction between additive theories and transformative theories, it seems to me, can be applied to theories of reasoning just as well as it fits theories of rationality.

According to Boyle (2012), most contemporary theories of mind endorse an additive view since they wish to emphasize the continuity between the human mind and the mind of non-human animals. Evolutionary accounts of mental development suggest that the human mind is, for the most part, highly similar to the mind of cognate mammalian species. If rationality is accepted as a distinguishing mark of the human mind at all, then the human mind is the result of – so to speak – simply adding the feature of rationality to the non-human mind. Yet, the emphasis on continuity and similarity between human and non-human mind is a rather recent development – at least if we take more than two thousand years of philosophical tradition into account. In this sense, the additive view is the new approach on the market, while

> [...] the [Aristotelian] tradition holds [that] the presence of rationality does not just add one more power to the human mind, or increase the scope and efficacy of mental powers already present in nonrational creatures. Rather, rationality transforms all of our principal mental powers, making our minds different *in kind*

[27] This distinction is already foreshadowed in Boyle (2012). I wish to thank Frank Hofmann for suggesting Matthew Boyle's texts to me.

> from the minds of nonrational animals. [Footnote omitted] (Boyle, 2012, p. 395)

Boyle (2012; 2016) presents interesting arguments against additive theories and provides noteworthy elaborations on more traditional accounts to rehabilitate the transformative view. There is no need to go into details regarding Boyle's attack against additive theories and his vindication of transformative theories here. The relevant point I am concerned with now is that the way I have set up the example where Chrysippus' dog and I share the same line of thought might suggest that I am committed to an additive view regarding reasoning. It might seem that I am out to claim that language, which enables us to check for correctness$_i$, is the special feature which yields a reasoner when added to a non-linguistic mind. Given the way I set up the discussion, it is certainly easy to get this impression. It might even be fair to say that I have invited this interpretation. But, in fact, I do not see any compelling reason to commit myself to an additive view. I prefer to stay neutral in this regard and therefore wish to make clear that my account is compatible with an additive view as well as with a transformative approach.

If a transformative view is correct, then my example with Chrysippus' dog on pp. 76 ff is, to say the least, misleading. Applied to my account of reasoning (instead of rationality as its proper place of application), the transformative theory arguably says that having a language affects all other aspects of the mind in a way that a reasoner cannot even share mental content with a non-linguistic creature. For Boyle (2012, pp. 399 f),

> [a] crucial implication of the Classical View [i.e., the origin of the transformative account] is that rationality is *not a particular power rational animals are equipped with, but their distinctive manner of having powers.*

As is evident from this quote, Boyle (2012; 2016), in his discussion of the transformative and additive views, is concerned with rationality and its effect on mental powers. I would translate this to the discussion at hand – i.e., the discussion about language and its effect on mental content – by saying that language is not a particular feature linguistic creatures are equipped

with, but their distinctive manner of thought, i.e., their distinctive manner of having and operating with mental content. This is by far not the only way to apply Boyle's (2012; 2016) distinction to the topic under consideration here. Because other plausible ways of adopting the contrast between transformative and additive theories exist and because the adaptation constitutes a considerable shift away from Boyle's focus (i.e., away from rationality to language/reasoning and mental content/representation), I am unsure whether Boyle would approve of how I make use of his distinction. But since the proximity of rationality and reasoning is obvious, I hope that Boyle would at least recognize *his* distinction – and not a completely different contrast which is merely presented in his terminology – in the present discussion even though he might deem it a misapplication of the contrast carved out between additive and transformative theories in Boyle (2012) and (2016).[28]

The illustration I have provided by making use of Chrysippus' dog does not go well with a transformative theory. Yet this is merely a defect of the particular way of illustrating the point in question. The relevant claim that the reasoner is ontologically prior to (instances of) reasoning is not in conflict with either the additive nor the transformative approach. The claim that any instantiation of reasoning depends on abilities of the reasoner – and therefore on the question of whether an individual line of thought appears in a mind capable of reasoning, i.e., a reasoner, instead of making the property of being a reasoner dependent on the presence of reasoning – is independent of the question of whether a transformative or an additive theory of the human, or rational, or linguistic (or whichever mental feature we might be interested in) mind should be preferred.

[28] A different aspect where Boyle and I are certainly in agreement is the clear commitment to the claim that the property of being a human being must not be confused with the property of being a reasoner (see p. 65). Boyle (2012, p. 401) makes a similar point when he says that "[...] the concept *rational animal* seems to be such that other species at least *could* fall under it." Swift's Houyhnhnms, the noble and perfectly rational horses Gulliver encounters during his fourth voyage, would make a vivid illustration of this point, "[...] since by hypothesis they would also be rational animals, 'rational animal' cannot be a complete characterization of what it is to be a man. [Footnote omitted]" (Boyle, 2012, p. 401)

3.4 Conclusion

With these results at hand, we can steer for a more convincing middle course to navigate between the Scylla of the *conduit metaphor* [29] and the Charybdis of *linguistic relativity*, thereby approximating a more adequate understanding of the relation between mind and language. A naïve conduit metaphor can be ruled out on the basis of *PRA* and the subsequent considerations since, if reasoning constitutively depends on language, it is not the case that language *merely* serves to communicate preformed mental content. In conclusion, we have at least a partial answer to the question of how thought and reasoning relate to language: Language is not necessary in order to think, but the difference between having and not having a language (together with the other core abilities from p. 75) amounts (at least) to the difference between being a reasoner and not being a reasoner. That said, it is a live option that not all reasoning necessarily needs to be carried out in language.

Regarding linguistic relativity – i.e., the other end of the spectrum, opposing the conduit metaphor – the current state of scientific investigation, based on several elaborately designed empirical experiments, strongly indicates that the difference in cognitive effect between different (human) languages is at best marginal (cf. McWhorter, 2014, pp. xiv, 84, and 106). But even if we have at best a marginal effect of language on thought, the question remains: Why settle for a middle ground between two extremes and not go the full way towards linguistic relativity? The following Part II of this investigation will provide an indirect and partial answer to this question before we take an entirely different perspective on the relation between mind and language in Part III.

[29] The picture that language merely serves to "translate" our (language-independently) premolded thoughts into a public language for communicative purposes; *viz.*, that every kind of thought is primary to and independent of language.

Part II

Are Language and Culture Inextricably Intertwined?

Linguistic Relativity

Chapter 4

The Case of Linguistic Relativity

4.1 Linguistic Relativity and Reasoning

We have already seen that, how, and why language is constitutive for reasoning. But language is not constitutive in the sense that it fixes the correctness conditions needed for reasoning. This would open the gates for the worst kind of linguistic relativity, since it might be the case that each language fixes different correctness conditions for reasoning. The correctness conditions for reasoning are, however, not dependent on any individual language and are therefore not language-relative. What language provides is not the correctness conditions themselves but *access* to correctness conditions which are fixed independently of any individual language. So, every language provides access to the very same correctness conditions although different languages may provide this access in slightly different ways – as will be shown with the aid of an obnoxiously sexist fallacy in a moment.

Although there is no difference among languages regarding which correctness conditions they provide access to, there may be differences regarding how this access is provided. Some correctness conditions might be provided more transparently in one language than in another. To illustrate such a possibility we can take a look at the following, very bad argument:

1. Only man is rational.

2. No woman is a man.

$\therefore$ 3. No woman is rational.

Although this argument is obviously fallacious, it is not as obvious which kind of fallacy it instantiates. It might be assessed as either formally valid – as an instance of *Modus Cesare* – with an ambiguous middle term, or as formally invalid for lacking a common middle term. I think that the categorization of equivocations as either formal or informal fallacies is somewhat arbitrary. I will, however, follow the standard account and treat this argument as an informal fallacy with a formally valid structure.

So, the fallacious argument in question turns on an equivocation of the term "man," which occurs with the same meaning as "human being" in the first premise and with the same meaning as "male (adult) human being" in the second premise. This equivocation cannot be reproduced, e.g., in German since there is no German expression for the middle term with a corresponding ambiguity. So the flaw in this reasoning is even more striking if we were to formulate the argument in German language rather than in English.

The message to take from this example is of course not that German is generally superior in the way it provides access to correctness conditions. The previous example was arbitrarily chosen, and a different example might have come out just the other way round by illuminating a flaw that is more readily detected or avoided in English than in German. The lesson to learn is that a line of reasoning is correct or incorrect in any language because the correctness or incorrectness of a line of reasoning is not dependent on the language which was used to express the reasoning in question. But different expressions of the same line of reasoning in different languages can sometimes highlight or obscure flaws in their respective formulations of the argument or line of reasoning in question.

In this sense, Whorf's (1956, p. 214) tentative suggestion to make oneself "[...] familiar with very many widely different linguistic systems" is reason-

able, but not because a monolingual individual is "[. . .] constrained to certain modes of [reasoning . . .]"[1] (Whorf, 1956, p. 214). The plausibility of Whorf's suggestion rather comes from the fact that certain flaws in reasoning might in some languages be more readily detected than in others – as in the example above where German does not permit the fatal equivocation because there is no German term available with a corresponding ambiguity. Being aware of such differences can certainly be helpful in some situations.

It might even sometimes be very helpful to translate a line of argument into another language in order to check whether an initially apparent cogency might only be due to a seductive formulation instead of the solid argumentative qualities of a line of reasoning. I presuppose, of course, that, if an idea or line of reasoning is cogent, it will be possible to find a convincing formulation in any language – as long as the language in question is complex and rich enough to represent the content to be expressed. That being said, if the *prima facie* persuasiveness gets "lost in translation," it is a good policy – I think – to reexamine a line of reasoning in another language and consider the possibility that it was a certain way of expressing things, rather than the content's quality to be evaluated, which lured our intellect.

The fundamental conviction which motivates such a position is, in a nutshell, that actual content can usually be translated without loss. Nuances which strike even experienced polyglots as persistently resisting satisfactory translation have a high chance of simply being irrelevant for the quality of reasoning on trial. What I have in mind here is, e.g., what Gottlob Frege (1960, p. 73) called "[. . .] the light in which [a] clause is placed [. . .]" without affecting its truth value, or the "[. . .] coloring (*Färbung*) and illumination (*Beleuchtung*) of the thought." (Mohanty et al., 1974, pp. 88 f)[2] I will not set out to argue for this view but merely leave it as an ideological backdrop – so to say – which might help to contextualize the rationale and motivation behind several points made in the vicinity of this section.

Nevertheless, a monolingual English speaker is of course able to detect the

[1] Whorf originally talks about interpretation, not reasoning.

[2] See also Frege (1956, p. 295): "What is called mood, fragrance, illumination in a poem, what is portrayed by cadence and rhythm, does not belong to the thought."

equivocation from p. 88 on her own without the need to ask a German speaker for help or take recourse to German language patterns. Consequentially, this example is far from suggesting any noteworthy restrictions in reasoning or conceptual systems due to one language in comparison with another. In consequence, the outcome that reasoning (constitutively) depends on language cannot be used to support claims of linguistic relativity.

4.2 The Wrong Way to Linguistic Relativity

Theories of linguistic relativity deserve a far more detailed discussion than can be provided here. I therefore do not claim that linguistic relativity can be refuted on the basis of these few remarks. Yet I am convinced that most accounts of linguistic relativity are deeply flawed and often radically overemphasize the allegedly all-pervading effect of different languages on our mental life. I mention only one problem many accounts of linguistic relativity suffer from, though not all of them (cf. Gumperz & Levinson, 1996, p. 1): Linguistic relativity theorists are frequently tempted to amalgamate language and culture in their notion of language. This happens sometimes implicitly; sometimes the mingling is explicitly endorsed as unavoidable.[3] I take this to be misguided and wholeheartedly subscribe to the statement that "[o]bviously, to even pose the question as to whether language and culture are related, there must be a sense in which the two can be distinguished." (Enfield et al., 2014, p. 13)

As a consequence of this amalgamated notion of language and culture, the linguistic relativist's claim "[. . .] that the particular language we speak influences the way we think about reality" (Lucy, 1997, p. 291) often amounts to the truism that our way of thinking is affected by our *linguistic-cum-cultural* situation. This should hardly come as a surprise, given that our cultural situation comprises all aspects of our education and upbringing. Theoreticians who use such an amalgamated notion of language therefore owe us an argument that their theories have anything to do with an actually

[3] See, e.g., Ahearn (2017, pp. xiii, xiv, 8, 20 f, 32, 56, 73, 92 f, 112, and 116).

linguistic impact on thought, independent of cultural effects in a broader sense.[4]

However, ignoring the intricacies of linguistic relativity here can be justified by pointing out that theories of linguistic relativity crucially differ in focus from the present investigation. I am concerned with the question of how thought, and especially reasoning, is dependent on (having a) language *per se*, i.e., what Lucy (1997, p. 292) calls the 'semiotic level.' Linguistic relativity, on the other hand, primarily makes claims about cognitive effects of speaking one language rather than another – Lucy's (1997, p. 292) 'structural level' – while it usually has little to say about differences in thought resulting from speaking any language in contrast to none at all (cf. Enfield et al., 2014, p. 8). These questions are sometimes not even clearly distinguished and consequently get confused in pertinent inquiries.[5] This makes the theoretical underpinning of alleged achievements in studies of linguistic relativity appear rather dim in many cases. Regarding this investigation, the question of how *different* languages affect thought is of secondary interest. The preconditions of thought and reasoning in general and the cognitive difference between linguistic and non-linguistic individuals are the central focus of the inquiry at hand.

Although linguistic relativity and my theory are logically independent of each other (*viz.*, the truth or falsity of one is of no direct consequence for the truth or falsity of the other), I wish to remain in the vicinity of linguistic

[4] This point is also repeatedly emphasized by McWhorter (cf. 2014, p. 139), who insists that the allegedly demonstrated effects of language-and-culture on thought come from culture (cf. McWhorter, 2014, pp. 12, 81 f, 103) or environment (cf. McWhorter, 2014, pp. 18 f), but not from language.

[5] It might also be the case that linguistic relativists frequently fail to properly distinguish between what David Lewis calls "[...] *languages*, [i.e.,] functions from strings of sounds or of marks to sets of possible worlds, semantic systems discussed in complete abstraction from human affairs, and what [he calls] *language*, [i.e.,] a form of rational, convention-governed human social activity" (Lewis, 1983, p. 166). The latter "[...] is regarded as part of human natural history [...]" (Lewis, 1983, p. 188), while the former "[...] are regarded as semantic systems [...]" (Lewis, 1983, p. 188). I am certainly concerned with an abstract system, rather than with a social activity, when I talk about language (*per se*). However, I am inclined to follow John Searle's terminology, rather than David Lewis's, where the roles of "language" and "languages" seem to be inverted, since for Searle (2011, p. 38) "[...] languages (as opposed to language) are conventional."

relativity accounts for a little longer. It should be clear by now that I think most of what haunts the discussion about language and mind under the label "linguistic relativity" is misguided. Still, I have not sufficiently argued for this claim. Due to the popularity of the linguistic relativity hypothesis, a wide variety of quite different claims with wildly varying grades of credibility aspire to be "the" linguistic relativity hypothesis.

Scope and intensity of linguistic impact on reasoning are quite diversely rated, according to different versions of the linguistic relativity hypothesis. Is thought determined or merely influenced by language? Do differences in lexicon affect thought or is the linguistic effect on thought restricted to very fundamental grammatical patterns of individual languages? Is there a linguistic impact on perception or does the language we speak merely affect how linguistically invariable perceptions are cognitively processed, *viz.*, does language merely affect cognition while leaving perception untouched? How these and many more questions are answered by different versions of linguistic relativity accounts needs to be differentiated before a proper assessment is possible.

I will neither assess different strands of the discussion about linguistic relativity, nor will I try to conclusively refute linguistic relativity theories. Instead, I will investigate what I take to be a common underlying assumption which motivates claims of linguistic relativity in many cases. This assumption, I conjecture, is also responsible for the popularity of the amalgamated notion of language-and-culture I mentioned on page 90. The assumption in question might be called

"the claim of the necessary sociocultural embeddedness of language."

I will show that this assumption, despite the fact that is seems to appear obviously true to many scholars, is false in most of its possible interpretations.[6]

[6] I will consider three specifications of the claim of the necessary sociocultural embeddedness of language:

(1) there cannot be a national language without its respective culture,

(1') every language needs to be embedded in its own culture

(which, I suppose, comes closest to what advocates of linguistic relativity feel sympathetic

Still, showing that the widely shared assumption in question is false does not amount to a refutation of the linguistic relativity hypothesis. As far as I can see, linguistic relativity theories do not require the truth of the assumption I will attack. Although the presumption in question and linguistic relativity are logically independent, I conjecture that they may not be psychologically independent. That is to say, while a refutation of the presumption in question does not prove the linguistic relativity hypothesis wrong, it might undermine an important motivational factor to endorse linguistic relativity in the first place. Still, this is admittedly quite speculative, and I will leave it to proponents of linguistic relativity to decide about the truth value of my conjecture about the psychological relevance of the presumption I will argue against. All the same, that the widely shared presumption in question is incorrect shall be shown in what follows.

4.3 Language and Cultural Embeddedness

The claim that "[...] language, thought, and culture are so intimately interwoven that to study any one of these is to study the other two as well" (Ahearn, 2017, p. 116) because "[...] disentangling language, thought, and culture from one another [...] turns out to be nearly impossible" (Ahearn, 2017, p. 112) is often either explicitly endorsed or simply (and silently) presupposed by many linguistic anthropologists and social scientists. I will focus on the claim that neither language nor its study can be separated from culture (and its study) and leave the corresponding questions relating to thought aside. The claim that language and thought cannot be studied separately from each other, or that to study language is (the best way) to study thought, occupies a special position since it is sometimes seen as the hallmark of analytical philosophy (cf., e.g., Dummett, 2014, p. 5) – a tradition I explicitly

to), and lastly

(2) there cannot be a language without being grounded in some culture, but not necessarily its own.

I merely mention these three interpretations now, since they will be properly introduced and discussed only at the very end of Part II on pp. 129 f.

identify with. Notwithstanding the prominent place this account takes, I do not endorse it and consequentially do not presuppose it. I also oppose the claim that this view correctly characterizes analytical philosophy, but this is a different story (to be told – I hope – on a different occasion).

As a starting point for investigation for numerable social scientists often serve claims of "the social nature of language" or "the socially embedded nature of language" (Ahearn, 2017, p. 21). Such claims about the *nature* of language have far-reaching consequences. Proponents of this claim might mistake the social embeddedness of language for an *essential* property of language, as in the particularly telling statement that "[l]inguistic anthropologists [. . .] maintain that the *essence* of language cannot be understood without reference to the particular social contexts in which it is used" (Ahearn, 2017, p. 8; emphasis added). Since this quote is taken from a current linguistic anthropology textbook, it can be regarded as being fairly representative of a widely shared approach to language.

This outlook, I claim, is mistaken since it suffers from severely overemphasizing the relevance of social context for a study of language in general. Although the role language plays in our social interactions is highly important and although this role is fascinating and far-reaching for us as language users, it is still an *accidental* feature of any language that it plays the role it plays – or any role at all – in the social world. Otherwise a language which is not used as a social means of communication would be impossible. I cannot, however, find a contradiction in the concept of a language which is not put to use in this way. In any case, the assumption that the nature of language requires an intimate connection with culture appears to be deeply rooted in several intellectual traditions. I take this view, and its role for the study of language, to be of some importance even independently of the argumentative main line of this investigation as a whole. So I will dedicate some room to probing this view.

4.3.1 Necessity or Essentiality?

Let us start by coming back to claims about the nature or essence of language as being socioculturally embedded. Since claims about the *essence* of language might result from loose talk, some clarification seems to be in order in regard to this point. We can explain the distinction between essential and necessary properties by following a classic example from Kit Fine (cf. 1994, pp. 4 f): If we consider Socrates and the set whose sole member is Socrates (let us call this set "singleton Socrates"), then it is necessary for Socrates to belong to singleton Socrates. Although it is plausible to assume that it is essential for singleton Socrates to contain Socrates as its only member (since sets are individuated by their members), it is hardly an essential property of Socrates to belong to singleton Socrates. The essence of Socrates – if there is such a thing – certainly has nothing to do with sets. So, although every essential property is a necessary property, it is not the case that every necessary property is also an essential property. In what follows, I will not be concerned with the stronger claim that language is essentially socioculturally embedded but, instead, only with the weaker claim that language is necessarily socioculturally embedded. It is the latter claim that shall be put under scrutiny in what follows, but every challenge to this weaker necessity claim will also be a challenge for the stronger essentiality claim.

Before we go on, one caveat needs to be inserted immediately: I grant that the term "essential feature" may be used differently from what I just sketched as the proper meaning of "essence." For example, in the Aristotelian tradition, an essential feature of, e.g., a kind need not be present in every specimen of the kind in question. Matthew Boyle (2012, p. 408) lucidly explains:

> On the Classical View [which builds on the Aristotelian tradition], propositions about *the essential features of human beings* are propositions about the kind *human being* itself, and there is no immediate inference to be drawn from such truths to free-standing propositions about what particular individuals of this kind are like.

If the notion of an essential feature is taken in this sense, the claim that cultural embeddedness is an essential feature of language cannot, of course, be refuted by citing "counterexamples" of languages which are not (necessarily) culturally embedded – which is, by the way, the overarching argumentative strategy I wish to realize in the following pages. Yet this Aristotelian or classical notion of an essential feature is not how the term "essential feature" is usually put to use nowadays.[7] Also, the quote from above proceeds as follows:

> This is not, of course, to suggest that truths about the kind and truths about individuals of that kind are simply unconnected: they are connected inasmuch as the truths about the kind describe how things go for individuals of that kind *if nothing interferes*. But to allow for the possibility of interference is to allow for the possibility of exceptions which do not disprove the rule. (Boyle, 2012, p. 408)

At first glance, this might seem to help the theoretician who wishes to insist that cultural embeddedness is an essential feature of language. If confronted with (apparent) counterexamples, it can still be claimed that "essential feature" needs to be understood in the Aristotelian sense, which allows exceptional exemplars that fail to exhibit the essential feature. Nonetheless, on closer inspection, it would arguably be a stretch to say that the examples I am about to discuss below (namely Klingon and Esperanto) would have been culturally embedded in the proper sense if only nothing interfered.

It might be possible to elaborate a cogent defense against the criticism which I will present shortly along these lines. This would amount to the following: any argument for the claim that language is not necessarily socioculturally embedded which is made in order to refute the claim that language is essentially socioculturally embedded is made in vain. For the latter claim – correctly understood, *viz.*, taken in its Aristotelian meaning – does not imply the former claim. This might be a tenable position and perhaps even

[7] For an enlightening discussion of the differences between modern and traditional understandings of essential properties, see Boyle (2012, pp. 406 ff).

a promising defense against what is to come. I am currently not in a position to judge whether this line of defense captures what someone who feels attracted by the position I criticize might have had in mind. Still, the burden of proof, I think, is on my opponent in this case. However, I will use the expression "essential feature/property" in the modern way sketched before, so that every essential feature is also a necessary feature but not *vice versa*. This is the way in which I will discuss the claim that language is necessarily or essentially socioculturally embedded or that it is the very nature of a language which ties it to its respective culture.

I will not discuss the question whether there can be culture without language in any detail. Since we have very credible experimental evidence for culture in at least a handful of non-human species (cf. Laland & Hoppitt, 2003) which plausibly do not have language, it seems highly probable that culture is not dependent on language. This tentative claim is based on a rather broad notion of culture, according to which "[c]ultures are those group-typical behavior patterns shared by members of a community that rely on socially learned and transmitted information." (Laland & Hoppitt, 2003, p. 151) Interestingly, given this wide definition of "culture," the best empirical evidence for culture in non-human animals is, according to Laland & Hoppitt (cf. 2003, p. 154), available for fish, followed by birds, followed by whales, followed by non-human primates.[8] This is probably not the order of non-human species one would anticipate when they are listed in regards to culture, but the significance of this ordering of evidence should not be overrated. Laland & Hoppitt (2003, p. 155) emphasize that "[t]he best evidence for culture is found in the species that are most amenable to experimental manipulation" and not necessarily in the species which are the best candidates for actually having culture. So, that the distribution of available evidence for culture in non-human species is quite surprising in no way contradicts the common opinion that non-human primates are nevertheless the most plausible candidates for culture in non-human animals (cf. Laland & Hoppitt, 2003,

[8] More detailed information regarding the claim that some of the most impressive findings suggesting animal culture come from studying fish, can be found in Laland & Janik (2006).

p. 155). But still it holds "[...] that for chimpanzees, as for other non-human primates, the hard evidence that their 'cultures' are socially learned is not yet there." (Laland & Hoppitt, 2003, p. 151) The hard empirical evidence, as mentioned before, proves culture only for "[...] humans plus a handful of species of birds, one or two whales, and two species of fish." (Laland & Hoppitt, 2003, p. 151)

However, language without culture is definitely possible, and therefore the social embeddedness of language is not a necessary feature of language, as will be shown in the pages to follow. So, the claims that "[...] language must not be studied in isolation from social practices or cultural meanings [...]" (Ahearn, 2017, p. 20) or "[...] that language, culture, and social relations are so thoroughly intertwined that they must be studied in connection with one another" (Ahearn, 2017, p. 32) must therefore be mistaken as well. At least, such claims are mistaken if they are meant to apply to every viable and useful form of study of language and not only to a perspective on language which already presupposes a focal interest in language when used in sociocultural contexts. This would leave the apparently strong and therefore quite interesting statements cited above utterly useless and trivial in the end.

In any case, I take it that everything which is not an essential feature of the object or phenomenon to be investigated can certainly be abstracted away in a fruitful investigation which still does justice to its topic. That taking into account the social and cultural surroundings of a language provides a richer picture is not to be denied. That said, the stronger claim that any investigation which abstracts away from these factors is deficient and unable to appropriately capture its object of study is clearly wrong. Such a perspective can only be the result of mistaking an accidental feature of language for an essential one or of an unwarranted and inappropriate extension of a singular research interest beyond its scope to the whole domain of investigation of a certain topic – i.e., a tunnel vision which dogmatically excludes everything out of sight, certainly not an accusation any respectable researcher would like to hear.

4.3.2 Disentangling Language and Culture

My argument to show that language is not necessarily culturally embedded is as simple as can be: Languages without cultural embedding are not only possible. There actually are languages without a corresponding culture, so it cannot be true that it is necessary for a language to be socioculturally embedded. This line of argumentation might seem trivial, but, since I have not encountered it in the literature yet, I deem it warranted to elaborate on this line of argument more than I probably would have done otherwise.

Examples of languages without appropriate sociocultural embedding are first and foremost so-called fictional languages such as Klingon (from Gene Roddenberry's *Star Trek*; cf. Stockwell, 2006, p. 9), Na'vi (from James Cameron's *Avatar*; cf. Zimmer, 2009), Elvish languages (e.g., Quenya and Sindarin from Tolkien's *The Lord of the Rings*; cf. Stockwell, 2006, p. 8) and many more. All these languages are full-blown languages, ready to be learned and used. I therefore dislike the label "fictional language" as it may suggest that we are confronted with a fictional, and therefore unreal, language. This would clearly be a misunderstanding since the languages listed before are real languages. One does not merely pretend that they exist as one may pretend in fiction that some alien species or mythical creatures exist.

The languages in question do actually exist. Since they did not grow out of what is usually conceived of as the natural way for languages to come into existence – but were instead invented in the context of fictional works – the subsumption under the label "fictional languages" is reasonable and widespread. The "normal" way for languages to emerge is, I suppose, a situation where a means of communication evolves by continuous codification into an organically grown linguistic body, which is passed on, modified, and developed from one generation to the next. This is not how invented languages usually come to be, so they do not fit our paradigmatic picture of a language – at least not in terms of origin. Still, it needs to be emphasized that these languages are *real* languages, *viz.*, they really exist. Only their context – i.e., their social embedding and the "cultures"[9] where these

[9] It will become clear in a moment why I put the term "cultures" in scare quotes here.

languages are spoken – is fictional.

That so-called fictional languages sometimes get adopted by a fan base and are spoken outside of their fictional context is beside the point here. Also that the fictional cultures are partly imitated and thereby get transferred to the real world is irrelevant in this context. The Klingon language and culture might serve as an example for both cases, i.e., the adoption of a language and of a "culture" by a fan base. Nevertheless, that real people speak fictional languages does not make the languages "more real" than they already were, and that real people appropriate fictional cultures does not make these cultures "less fictional." Also, in order for a language to be adopted by its fan base, the language needs to exist already before it is applied in the real world.

An important distinction needs to be drawn among the so-called fictional languages before we continue. On the one hand there are really existing and actually invented languages coming from fiction. The examples I mentioned before – Klingon, Sindarin, Na'vi, etc. – fall in this category. The relevant feature of these languages is that they are fully developed languages ready to be learned, so to speak. I will call these languages "languages from fiction." On the other hand we have merely fictitious languages such as the Adamic language or the language of Atlantis. Given that there was never a divine language spoken before the Tower of Babel or an island of Atlantis, these languages are purely mythical and rightly deserve to be called "fictional languages." In order to avoid possible confusions, I will use the term "invented languages" to, *inter alia*, cover really existing languages from fiction, but not fictitious languages. So, not all fictional languages are invented languages: Merely mythical languages and languages from fiction are often thrown together under the label "fictional languages," but only the latter are also invented languages. The former languages do not count as invented languages because they were, strictly speaking, never invented. There is only a pretension to their existence, but in fact they do not exist. Something that was invented at a certain point in time, in contrast, clearly exists. Note, however, that the notion of invented languages has a wider extension than the notion of fictional languages, since there are invented languages which

were not created to serve the purpose of fiction.

A short terminological digression seems to be in order at this point. The label "invented language" has certain disadvantages as well. This designation might reinforce a misleading connotation coming from the conceived difference between artificially created languages on the one hand, and naturally originated languages on the other hand. While this contrast is not illusory, it may suggest that so-called natural languages were not invented (by humans) but simply popped into existence, "fell from the sky," or were given to us by God. I take it as a given that natural languages were also created and therefore invented by human beings. They were just not invented in the intentional, purposeful, and planned manner in which came about those languages which we are used to calling invented languages.

The term "planned languages" is sometimes proposed as an alternative for "invented languages." Yet this expression comes with an analogue problem: Most national languages (which are the paradigm case for natural languages) are acutely regulated and should therefore not be considered as growing in an "uncontrolled" manner. That these languages were not planned from scratch does not make them "unplanned." Overemphasizing a simplistic disanalogy between so-called natural languages on the one hand and so-called artificial languages on the other hand looms in the background of many terminological imperfections in the vicinity of the topic at hand. "Auxiliary language" (sometimes preceded by the addition "international") as a term to cover languages which are invented from scratch (usually to facilitate international comprehension) but are not fictional languages is even worse because this expression suggests that the languages falling under this label are not full-fledged languages or that they are inferior languages in some way. This is certainly mistaken. I am not aware of any expression to make use of here without any downside. So, I suppose, the perfect term which would allow me to avoid this side remark does not exist. However, as long as none of the misunderstandings mentioned before comes up, the terminology does not really matter. Every expression which is used in discussions about invented languages has its pros and cons. I will primarily use the term "invented languages" in what follows, but several alternatives might have served my

purposes just as well – given that the misleading connotations are ironed out.

I will treat examples of languages which are invented but cannot be called "fictional" later, especially in section 5.2 on pp. 115 ff. Before that, a possible rejoinder from the proponent of the necessary social embeddedness of language regarding languages from fiction deserves attention: Could the point I made so far about languages from fiction (e.g., Klingon and Sindarin) not be taken as a straightforward objection to my claim that language is not necessarily socially embedded? Does the fact that all languages from fiction seem to be always invented together with a cultural context not prove that a language without cultural context would amount to a hardly conceivable monstrosity? At least those languages I mentioned as examples of languages from fiction – Klingon, Na'vi, and Elvish languages – have their own fictional cultures. I do not know whether there might be a fictional language which was created without its own peculiar, fictional cultural embedding. That said, for the sake of argument, let us for the moment assume that at least every language from fiction (if not even every fictional language) was not only actually invented together with its own fictional cultural surrounding but also needed to be created with a cultural embedding (for whatever reason).

If we came across a fictional language which does not obviously have its own fictional culture, then we might say that the language in question is fictionally spoken in some actually existing culture. This would be the case, for example, in hypothetical accounts of alternate histories where we – i.e., actually living people in the actual world – came to live just as we do, with the slight difference that we do not speak the language we actually speak but, instead, a fictional language. Similar considerations regarding fictional scenarios should do, I hope, to accommodate all relevant varieties of the pertinent possible objection.

Does the at least *prima facie* plausible impression that fictional languages always come accompanied by their own fictional cultural embedding not reinforce the claim that a language cannot be separated from its culture? In the end it might seem that – as could be plausibly said of languages from fiction – if a language has no cultural context, the need for a language to be culturally

embedded is so strong that the culture needs to be invented in the process of inventing the language. So, the sociolinguist or the linguistic anthropologist might claim, what I have done up until now is by no means suited to disprove the claim that language and culture cannot be separated. On the contrary, I have just provided an additional reason to stick to the thesis that any alleged distinction between language and culture is futile, by pointing out that even languages from fiction (and probably also other fictional languages) – which arguably are not socioculturally embedded – need to be invented with their own (fiction of a) cultural surrounding. What could be a stronger case for the need a language has for a culture than the additional effort of language creators to invent a culture for their languages?

To think in this way would amount, I think, to committing what Saul Kripke (cf. 2011b, pp. 345 f) calls the *toy duck fallacy*. Since this fallacy is not widely known yet, even among philosophers, a few words of explanation are in order. I will clarify the toy duck fallacy and how it applies to the question at hand in the following section. Before we come to that, I wish to add one further point of clarification: I said in the previous paragraph that fictional languages, including languages from fiction, are arguably not socioculturally embedded. Before I can provide an argument for this claim, the toy duck fallacy needs to be explained. There is, however, an independent possible confusion, looming in the background, which should be pointed out now.

In a certain sense, it is correct to say that even invented languages, including languages from fiction, are socioculturally embedded because their inventors are. Still, this does not give us the right kind of sociocultural embeddedness. The fact that the inventor of a language is socioculturally embedded in a certain way – since she is part of a certain culture – is irrelevant for the claim that every language needs to have its *own* culture. This formulation should be taken in a lenient understanding which permits that it is still true that every language has its very own culture even if the cultures of different languages might coincide. That is to say, if we were to count different languages, e.g., Sindarin and Quenya, as being embedded in

the very same fictional Elvish culture,[10] this would not count as an objection against the claim that every language needs to have its very own culture.

However, in a certain way the culture of a language inventor is of course the culture which gave rise to an invented language. All the same, this culture clearly is not the culture of the invented language. The British culture is not the culture of any Elvish language although the inventor of the Elvish languages – J. R. R. Tolkien – was British. Likewise, the American culture is not the culture of the Klingon language although its creator – Marc Okrand – is American. The point I wish to make clear is that claiming that even invented languages necessarily have a culture because, in order for them to be invented, it takes a culture which yields a language creator is beside the point.

I assume that merely a clear statement – which I hope to have provided – of the fact that the culture of a language creator cannot automatically count as the culture of the language created – even without a proper and independent argument for this claim – can help to avoid some confusions which may have interfered in the following discussion otherwise. I will come back to this point a little later, but it is now time to present the toy duck fallacy.

[10] I am not sufficiently familiar with the details of Tolkien's work to judge whether this is actually a good example to illustrate my point. For the present purpose it should be fine as long as the example can be accepted as an at least possible way of describing the cultural and linguistic situation in Middle-earth. In any case, a friendly connoisseur backs up the scenario I sketched as a plausible description which comes close enough to the canonical interpretation, at least if we understand "culture" in a broad sense.

Chapter 5

The Counterexamples

5.1 Toy Ducks and Privative Modifiers

5.1.1 The Toy Duck Fallacy

Saul Kripke (2011b, p. 345) introduces the toy duck fallacy with the following example:

> A parent takes a child to a toy store. The toys are plastic models of various animals. The child points to a toy duck and asks, "Is that a goose?" The parent responds, "No, that's a duck." (Blumberg & Holguín, 2018, p. 2058)[1]

With recourse to this example, Kripke forcefully warns us not to draw the wrong conclusions from the blatant naturalness and undisputed acceptability of the parent's response. Especially, we must not take the parent's response "No, that's a duck" – when uttered about the salient toy duck – to be a literal

[1] I take this depiction from Blumberg & Holguín (2018, p. 2058), but it is virtually identical to the relevant passage in Kripke (2011b, p. 345), except for the slight precisification that the child points to a toy duck when asking its question, which makes the chosen formulation somewhat more suitable for the present exposition.

However, for the sake of completeness, here is Kripke's original formulation:

> Suppose a parent takes a child to a toy store. The toys are plastic models of various animals. The child asks, "Is that a goose?" The parent says, "No, that's a duck." (Kripke, 2011b, p. 345)

truth. A toy duck is not a duck. It is a toy, not a duck; it is an artifact, not an animal belonging to the Anatidae family. Therefore, it is incorrect to count a toy duck among the ducks, notwithstanding the feeling of correctness we might have when confronted with the parent's reply. To think that the acceptability of the parent's response would license the conclusion that there are different kinds of ducks – real ducks on the one hand and toy ducks on the other hand – would be an error. This kind of inference would be a toy duck fallacy.

The application of the toy duck fallacy to fictional cultures is straightforward. That we can quite naturally talk about fictional cultures like the Klingon or Elvish cultures should not tempt us to think that we can sort cultures into the following categories: real cultures on the one hand and fictional cultures on the other hand. Fictional cultures are not cultures just as toy ducks are not ducks. So long as we keep the relevant facts in mind – namely that toy ducks are not a kind of duck and that fictional cultures are not a kind of culture – our loose talk in this regard is unproblematic. It is unnecessary to point out to the parent in Kripke's example that his classification is faulty since nobody would conclude that the biological typology of ducks needs to be expanded. As long as an analogous mistake is not triggered by loose talk about Klingon or Elvish cultures, it can be readily tolerated. That said, the danger of drawing the wrong conclusions is certainly more prevalent in the context of cultures than in the context of ducks. Therefore, increased mindfulness is certainly more important when we talk about fictional cultures than when we talk about toy ducks.

5.1.2 Privative Modifiers

A different way of fleshing out the error in question is to point out that "toy," as it is used in the expression "toy duck," is a *privative* modifier.[2] What

[2] I use the term "modifier" to cover various kinds of different privative expressions which belong to different word classes, such as adjectives, adjectivally used nouns, certain prefixes, etc. Since differentiating between those would be a superfluous complication of the discussion at hand, the neutral expression "modifier" is used instead of the more common term "privative adjective."

makes a modifier privative is its "[...] entailing the negation of the noun property" (Partee, 2007, p. 151; and Partee, 2010, pp. 275 and 279) when combined with a noun. Examples of privative modifiers are "counterfeit," "fake," "fictitious" (cf. Partee, 2007, pp. 150, 153, and 155; and Partee, 2010, pp. 274 f, and 277), as well as "former" and "past" (cf. Partee, 2007, p. 149; and Partee, 2010, p. 275), "spurious" and "imaginary" (cf. Partee, 2007, p. 155; and Partee, 2010, p. 279), and finally also "toy" (cf. Partee, 2007, p. 153; and Partee, 2010, p. 277). The reasoning behind privative modifiers is that, e.g., "a *fake gun* is not a *gun*" (Partee, 2010, p. 277), just as a *toy* duck is not a *duck*. Partee (2010, p. 277) speaks about "[...] a 'negative' meaning postulate", which can be depicted in the following way:

$$\forall x \, \forall P \; [pm(P)(x) \rightarrow \neg P(x)]$$

(cf. Partee, 2007, p. 149), where *"P"* represents any noun term, and *"pm"* stands for any privative modifier, such as the previously listed examples, or the prefix "non-", if grammatically applicable. A non-duck is clearly and most straightforwardly *not* a duck, and so is a toy duck, given that "toy" is also a privative modifier.

It should be noted that not all of Barbara Partee's examples of privative terms are indisputable. Two cases in point are "former" and "past." Partee (2010, p. 282) presents the following case:

> In English there does not seem to be any difference between *past* and *former* with respect to privativity – both are normally regarded as privative, although this is sometimes called into question: if a former/past senator is re-elected after spending some time out of office, is she then both a former/past and current senator, or does she cease being a former/past senator?

This example regarding "former" and "past" is, however, not very convincing in my opinion. I conjecture that the dubitable privativity of "former" and "past" in the case of the former/past senator can be explained by pointing out the difference between attributive and predicative uses of adjectives.[3]

[3] Not to be confused with Keith Donnellan's (1966) distinction between attributively

The difference is the following: If an adjective is used predicatively, then the adjective straightforwardly applies to the thing the adjective was (truthfully) applied to. For example, a green car is a car, and it is green. Here the adjective "green" is used predicatively, and therefore the car in question falls under the extension of "green" if it was correctly addressed as a green car. The situation is different with attributively used adjectives. An old friend, for example, is not necessarily a friend and also old. An old friend might be fairly young but long-known. Therefore, although it is correct to address someone as an old friend, the person in question does not necessarily fall under the extension of the adjective "old." The friend in question might, of course, have a high age, but she does not have to, in order to be an old friend.

This significant difference between attributively used and predicatively used adjectives can serve to explain the recalcitrant data quoted by Partee in the following way. As far as adjectives are concerned, we can only speak of privativity if they are used predicatively. In the case of the former/past senator, "former" or "past" are clearly used attributively since it is not the case that the person in question is/was a senator and also former or past. Attributively used predicates may always fall under a category which Partee (2010, p. 276) calls "plain" nonsubsective adjectives: "The 'plain' nonsubsective adjectives (*alleged*, *possible*) have no meaning postulate; this class is 'noncommittal': an *alleged murderer* may or may not be a *murderer*." Also in this example we have an attributively used adjective since, even if the person in question is a murderer, she is not alleged and also a murderer. So far, the claim that certain adjectives are privative *if they are used predicatively* (this is the precisification I suggest for the question regarding privative adjectives) cannot be called into question by citing examples where the privativity of adjectives is questionable when they are used attributively.

However, Partee is not done yet. She continues:

> There are certainly other unclear cases in English: witness the

and referentially used definite descriptions. Attributive uses in Donnellan's sense are characterized by the fact that the thing designated by a definite description fulfills the predicate used to single out the designated object. This stands in marked contrast to attributively used adjectives, as will become clear in a minute.

> uncertainty in classifying *retired*, *dead* as intersective versus privative. Is a retired professor a professor? Probably yes. Is a retired CEO a CEO? Probably no. (Partee, 2010, p. 282)

Here we have a clear case of a predicatively used adjective since a retired professor is (probably) a professor and is (certainly) retired. Even if a retired CEO is (probably) not a CEO anymore, she still is retired. This is good evidence that "retired" is not privative. We can make a similarly clear case against the privativity of "past" by pointing out that a past event is past and also is/was an event. This might settle a question which cannot be conclusively answered by Partee's spurious example with an attributively used instance of "past."

More importantly, Partee (2007; 2010) does not only wish to cast doubt on *some* allegedly privative adjectives. Rather, she claims "[...] that no adjectives are actually privative." (Partee, 2010, p. 279) Still, given that she puts so much argumentative weight on sentences like "Is that gun real or fake?" (Partee, 2010, p. 274) to motivate her "no privatives" hypothesis (Partee, 2010, p. 283), it is a plausible presumption that Partee fell prey to a toy duck fallacy herself. At least I am inclined to consider the fake-gun-real-gun and other examples she provides to be toy duck cases.[4]

[4] Partee (2010, p. 277) writes that "[o]ne nagging problem is the evident tension between the apparent truth of [1] and the undeniable well-formedness and interpretability of [2].

[1] A fake gun is not a gun.

[2] Is that gun real or fake?"

The fact remains that the truth of [1] only seems to be in tension with the blatant acceptability of [2] if we are tempted to conclude from the naturalness of the question "Is that gun real or fake?" that it is (literally) correct to call a fake or toy gun a gun. Yet drawing the conclusion that a toy or fake gun is a gun on the basis of this consideration – namely that the question "Is that gun real of fake?" is not defective – is simply committing the toy duck fallacy. I therefore claim that Partee's (2010, pp. 279 ff) argumentation for the claim that there are no privative adjectives – her "no privatives" hypothesis – is faulty for committing the toy duck fallacy (and/or overreacting on the basis of misleading evidence arising from falling prey to the toy duck fallacy). If we are on our guard against the toy duck fallacy, we can see that there is no nagging problem (indeed, no real problem at all), because there is merely an *apparent* tension between the *evident* truth of [1] and the naturalness of [2]. (The attentive reader has probably realized that, in the previous sentence, I merely shifted the italicized expressions around in the quote from Partee (2010, p. 277) which opened this footnote.) So, I claim, mere awareness of the toy duck fallacy defuses

5.1.3 Fictional Entities and Toy Ducks

Let us bring the discussion back to fiction. The most relevant application of the toy duck fallacy for our purpose at hand is to be found in Kripke (2013), where an analogy is drawn between real ducks and toy ducks on the one hand and real people and fictional characters on the other. Since the ontology of fictional entities is more intricate than the ontology of toy ducks, some clarifications are needed before we can proceed with an application of the toy duck fallacy to cases of fiction.

I agree with Kripke (cf. 2013, pp. 80 f) that to say of an entity, e.g., Hamlet or Sherlock Holmes, that it is not real or that it is merely fictional does not mean that the entity in question does not exist. Hamlet and Sherlock Holmes clearly do exist, and we can state many true things about them. We can, for example, answer questions about when they were created and by whom. We can say that they are very popular, that many literary theorists have written countless pages about these characters, and so on. It is also clear that neither Hamlet nor Sherlock Holmes are real since they are fictional characters. The confusion Kripke wishes to resolve is that, just because something is not real, because it is fictional, this does not mean that it does not exist. How could Hamlet and Sherlock Holmes not exist? They were created at a certain point in history, and, since the time when they were created, they exist.

So, here is one analogy we can draw between toy ducks and fictional characters: Toy ducks are not real ducks just as fictional people are not real people. Nonetheless, the fact that toy ducks are not real ducks does not make them less real, *viz.*, it does not make them exist to a somehow diminished degree or in an inferior way. The same holds true for fictional characters as well. The fact that fictional persons are not real persons does not mean that fictional persons do not really exist. They do, but they are not persons. Even though Sherlock Holmes and Hamlet are real persons *according to their respective stories*, they are not persons in reality. In reality, they are fictional persons. Instead of diving deeper into the ontology of fictional entities, we

Partee's (2010) alleged problem cases, and we have no reason to abandon the view that expressions such as "toy," "fake," and other expressions – including, notably also "fictitious" and "fictional" – are indeed privative (if used predicatively).

should finally come to the application of the toy duck fallacy in regard to fictional entities.

We know that a toy duck is not a duck because "toy" is a privative modifier. To commit the domestic version of the toy duck fallacy, so to say, amounts to being fooled into believing that a toy duck is a duck because it feels quite natural to address a toy duck as a duck as Kripke's toy duck story, introduced on page 105, strikingly testifies. The toy duck fallacy, if committed in regard to fictional entities, can fool someone into believing that a fictional person is a person because it also feels quite natural to say that Sherlock Holmes and Hamlet are persons. However, "fictional" is a privative modifier just as "toy" is. Therefore, Hamlet is not a person since "[. . .] there is no such *person* as Hamlet [. . .]" (Kripke, 2013, p. 148) although there is the fictional person Hamlet. I repeat: A fictional person is no more a person than a toy duck is a duck.

The attentive reader has probably noticed that "fictional" is not included in Barbara Partee's list of privative modifiers, assembled on page 107, although "fictitious" and "imaginary" appear there. There are important similarities, as well as differences, between "fictitious" and "imaginary" on the one hand and "fictional" on the other hand. When we call something fictitious or imaginary, we often mean to say that it does not exist at all. The birthday of my fictitious aunt, for example, which I may use as a pretense to justify my absence from a social event I would rather like to avoid participating in (not that I would ever do something like this), is completely made up. Neither the aunt nor her birthday exist in any ontologically robust way.[5] They are

[5] By "ontologically robust" I mean: over and above a mere intentional existence, i.e., an existence only as a representation in my mind without anything which corresponds to this representation. This is a very rough and tentative characterization since writing my fake excuse down will create a representation of my fictitious aunt which does not exist in my mind. All the same, an externalization of this kind will make neither my fictitious aunt nor her birthday one iota more real. In the case of Hamlet and Sherlock Holmes, on the other hand, writing down their stories is an integral part of bringing them into existence. The details of an ontological theory of fictional characters are quite intricate and will not be pursued any further here. I hope that, even without more elaboration on the metaphysics of fiction and fictional entities, it will be clear enough in which sense a fictitious character – in contrast to a fictional character – does not exist in an ontologically robust way. For our purposes, however, it will be sufficient to say that fictitious entities do not exist at all – *viz.*, that they are not real –, at any rate.

completely imaginary. Calling something "imaginary" or "fictitious" does not always license a conclusion to plain non-existence. The fictitious excuse I sketched before is just a pretense, and, given that "fictitious" is privative, it cannot count as a (justified) excuse at all. But the pretextual reason for not attending the event in question is still there. It just is no reason; it is merely a pretense.

So much for the case of "fictitious." But how about "fictional"? The important difference between "fictitious" and "fictional" is that "fictional" *never* warrants an inference to plain non-existence while "fictitious" at least sometimes does, or so I claim. If something is fictional in the sense of appearing in a work of fiction, then the fictional entity exists. It was invented by an author, may become part of a common cultural heritage, and can, e.g., be studied by literary scholars. Still, "fictional" – and this is the important similarity mentioned before – is just as privative as "imaginary" and "fictitious" are. Therefore, if something is a fictional X – be it a person, a place, a country, a law, an event, or what have you – then it is not an X at all although, as emphasized before, it is still something – *viz.*, it exists; but it is not an X. It must be kept in mind that the question of the legitimacy of an inference to non-existence is completely independent of the question whether a term is a privative modifier. Privative modifiers warrant an inference to the fact that things are not of a certain kind, but they do not – at least not *qua* being privative modifiers – license the inference that things do not exist.

5.1.4 Klingon Culture and Language

Returning to fictional cultures, finally, we must conclude that a fictional culture is not a culture at all. There is no Klingon culture, for example, although there is a fictional culture called "Klingon" in *Star Trek*. This fictional culture exists since it was invented at a certain point in time by Gene Roddenberry, I gather, who created the *Star Trek* franchise. Although Klingon is a fictional culture, it is not a culture. But how about the Klingon language? There actually is a Klingon language. It was initially invented by

the American linguist Marc Okrand for the *Star Trek* franchise and eventually grew beyond the screen. It is estimated "[...] that about 20 or so people in the world have a high enough level of proficiency to hold a conversation purely in Klingon. But many more get by." (Prisco, 2018) So, given that Klingon is an actual language and that there is no Klingon culture (although there is a fictional Klingon culture) it cannot be true that a language necessarily comes with a culture. For languages like Klingon and most other languages constructed for, or by, the entertainment industry – taken in a wide sense – it holds true that there is no culture correlated with these languages at all even though there is usually a fictional culture where the language is (fictionally) "embedded."

For the still unconvinced reader who might be tempted to think that at least a fictional culture needs to be in place for a (constructed) language to work, here is an interesting quote from Marc Okrand, the creator of the Klingon language himself, which strongly suggests that the Klingon language preceded most of the fictional Klingon culture. I take it, by the way, for granted that the Klingon language had come into existence at the latest by 1985 when Okrand (1992) was first published. Describing how he came to write a book (namely Okrand's *The Klingon Dictionary*) which explains how the Klingon language works, including its grammar and a dictionary, Okrand says:

> That [i.e., coming up with a Klingon-English bilingual dictionary] was actually harder than describing the grammar, because I had to decide what words to actually invent. I decided to not make up any words having anything to do with Klingon geography or Klingon culture. I know it sounds strange to have a dictionary about Klingon that doesn't deal with that aspect, but the reason is that I'm not a writer, I don't write the stories or the movies and I didn't want to make something up that down the road would turn out wrong because of a TV episode or a movie. So I would let writers make up the culture, and come back afterward to say "This is how you call it." Not the other way around. (Prisco, 2018)

Also, the introduction to Okrand's *Klingon Dictionary* comes with a disclaimer that words relating to Klingon culture – e.g., "[...] words for native tools, customs, [...] and vocabulary dealing with food" (Okrand, 1992, p. 10) – are, for the most part, not covered. This is not "[b]ecause research is not yet completed [...]",[6] as Okrand (1992, p. 9) claims but because nobody had made up these parts of Klingon culture yet, despite the fact that many surprisingly detailed and credible "facts" about Klingon sociology are presented in the same place (cf. Okrand, 1992, pp. 9-12). Any attempt to speculate that a construction with enormous gaps like these in its vocabulary cannot be considered a real or full-blown language, I think, cannot withstand even mild scrutiny. Was English incomplete (and therefore not really a language) before there was a word for cyber-bullying? Certainly not. But which aspects of (human!?) culture need to be covered before we can call a system of communication a language? I hope that this question will strike at least most readers as sufficiently dull to not even deserve any serious consideration. I contend that all kinds of attempts to hedge, in order to generally exclude constructed languages from one's notion of a language proper, are doomed to fail from the outset for a simple reason: many constructed languages simply are full-blown languages proper, so an argument set out to prove the opposite must simply be flawed.

However bullet-proof I take this argumentation to be, I am almost certain that some readers will remain unconvinced because they suspect that my categorical distinction between fictional culture on the one hand and culture on the other hand is forced, faulty, or spurious. I do not believe that this is the case, but I take the point. There is another way to demonstrate that cultural embeddedness is neither an essential nor a necessary feature of language because language without culture is not only a possible but an actual phenomenon. The following, second way of fleshing out this argument is entirely independent of the distinction between fictional and real culture. In order to demonstrate the case, we will leave "Hollywood languages" like

[6] Note that the whole book, i.e., Okrand (1992), comes under the pretense of being a study, carried out by some scientific research institute inside the *Star Trek* universe. A very nice example of a fictional context for a real book.

Klingon and kindred languages from fiction behind and focus instead on a different kind of constructed languages, the most prominent example of which is Esperanto.

5.2 The Argument from Conlangs: Esperanto

5.2.1 Introduction

The expression "conlang" is a portmanteau word from "constructed language."[7] A clear-cut distinction between languages from fiction on the one hand and conlangs on the other hand can certainly not be drawn by merely taking into account the structural and linguistic features – or, in other words, the intrinsic features – of the languages in question. The most plausible way to distinguish languages from fiction and conlangs is that languages from fiction are intimately associated with a fictional context, paradigmatically products of the entertainment industry and popular culture, notably fantasy and science fiction. Conlangs, in contrast, usually do not have any seminal connection with fictional works, but tend to be firmly rooted in the real world instead of being rooted in a fictional context.

The intentions of language inventors in the area of conlangs are as diverse as their projects. Some aspired to further peace and harmony among mankind, others wished to provide means to improve the functioning of the human intellect, and still others might simply be language aficionados and enthusiasts who devise language projects to indulge in a somewhat nerdy – or brainy, as we might rather say nowadays – kind of creativity. Add nearly every shade in between and mixture of these archetypical examples, and the resulting typology will still provide only a very rough overview.[8] So, the field

[7] The expressions "constructed language" and "invented language" should not be considered to be perfectly synonymous. A certain language variation might, for example, due to strict regulation, be legitimately considered to be constructed although it would probably seem odd to say that the variation in question is invented. However, such considerations will, for most people, probably stand on rather shaky intuitions. "Constructed language" and "invented language" can, at least for the purpose at hand, probably be used interchangeably in most instances.

[8] An excellent general overview of constructed languages – including also some lan-

of conlangs and their creators is manifestly a highly heterogeneous affair. For present purposes, however, only one constructed language is of further interest. We will take a somewhat more detailed look at the undeniably most prominent of all constructed languages, Esperanto, and at the customs of its language community to make a case for the existence of a dedicated Esperanto culture.

5.2.2 Some Background Information

The language Esperanto was created by the Polish ophthalmologist Ludwik Lejzer Zamenhof – or, equivalently, Ludwig Lazarus Zamenhof – in the late 19th century to aid international communication. Although Esperanto derives most of its vocabulary from other languages (most notably Romanic, Germanic, and Slavic languages, in that order; see also Wells, 2009, p. 376), it is undoubtedly a self-standing language. If the status of Esperanto as a language on its own was doubted on these grounds, then Romanic languages like French, Spanish, or Portuguese should also not be counted as independent languages since they derive most of their vocabularies from Latin. If it goes without question that Portuguese, French, and Spanish are distinct languages, then neither should there be any doubt about Esperanto in this regard.

The grammar of Esperanto is, unlike its vocabulary, not constructed in analogy with paradigmatic European languages.[9] In terms of linguistic typology, Esperanto should be considered to be an *agglutinative* language, in contrast to the class of inflected languages which includes most of the stereotypical European languages. Consider some examples for illustration: German, French, Portuguese, and Spanish are usually counted as highly inflected languages. Examples of European agglutinative languages are Finnish and Hungarian (cf. Bodmer, 1946, p. 196), as well as Turkish (cf. Mair, 2015, p. 67), but agglutinative language structures are far more common in the rest of the world, outside of Europe. The grammatical structure of Esperanto

guages from fiction, and in some cases also information about their creators – provide, e.g., Okrent (2009) and, famously, Eco (1995).

[9] But see Bartlett (2009, p. 76) for an apparently diverging opinion on this matter.

is, however, quite distinctive in its formal strictness and simplicity, and cannot be considered as being derived from any single natural language even though its structure is certainly inspired by already previously existing linguistic structures from several languages. Agglutinative languages are also sometimes called

> [...] agglutinating languages, [... and w]hat is most characteristic of such languages is that each affix, like an independent word, has a *distinctive* meaning. So derivatives [...] of an agglutinating language when classified according to case, mood, etc., have clear-cut uses, and the method of forming them is also clear-cut. Neither the use nor the form of derivatives described by the same name admits the perplexing irregularities of a typically *amalgamating* [i.e., inflected] language such as Latin, Greek, or Sanskrit. (Bodmer, 1946, p. 197)

Bodmer (1946, pp. 198 f) further explains that

> [...] two essential features are common to all the [agglutinating languages]. One is *great regularity of the prevailing pattern* of derivatives. The other is *comparative freedom from arbitrary affixes* which contribute nothing to the meaning of a statement. Thus grammatical gender [...] is completely absent.

Both of these features perfectly describe Esperanto. In general, the characterization of a language as an agglutinative language is a morphological categorization, i.e., based on the level of word structure, in contrast to, e.g., sentence structure. So, in languages of the agglutinating type, "[w]ords are formed by a root and a clearly detachable sequence of affixes, each of them expressing a separate item of meaning." (Iacobini, 2006, p. 280) Since "[e]ach affix carries only one meaning [...,] the semantic structure is directly reflected in the morphological articulation of the word [...]" (Iacobini, 2006, p. 280). This leaves us with "[n]o inflectional classes, [and, as already noted above,] no gender distinction." (Iacobini, 2006, p. 280)

It is comprehensible that these features attracted theorists of language, as well as language creators. The realization that there are languages which

work perfectly well without incorporating complex or even overcomplicated aspects, such as grammatical gender [10] and declination,[11] must have been revealing for those who were interested in the structure and functioning of language and who were primarily confronted with inflected languages due to their humanistic education. Linguistic aspects which the student of a language must tediously learn by heart in order to speak correctly turn out to be dispensable in other languages. In this situation, it is no wonder that

> [t]he veteran philologist Jacob Grimm first emphasized the merits of Magyar [i.e., Hungarian] and commended it as a model to people interested in language planning. The existence of such regularity in natural languages has left a strong impress on projects for a constructed world auxiliary. (Bodmer, 1946, p. 200)

Somewhat tendentiously, but not incorrectly, Bodmer (1946, p. 202) also remarks that "[t]he grammar of an agglutinating language such as Finnish (or Esperanto) is mainly concerned with meaning. The grammar of an amalgamating language such as Latin is mainly concerned with social ritual." In the face of these merits, it is reasonable that several constructed languages fall into the category of agglutinative languages. Among those already mentioned here are, e.g., Klingon and Quenya.

5.2.3 The Language of Esperanto

Although Esperanto clearly is a constructed language, its status as an actual, real, and full-blown language is hardly ever disputed – and if so, on very shaky grounds (cf. Bartlett, 2009, pp. 75 f; as well as Wells, 2009, p. 375). Esperanto is unfortunately often categorized as an "artificial" language, because it was invented from scratch – at least if we ignore the fact that virtually all word roots were taken from or inspired by existing languages and are therefore not really invented. However, I would prefer to avoid the expression

[10] Who ever tried to learn an inflected language will have realized how arbitrary or even bizarre grammatical gender often seems.

[11] Hardly any student of Latin probably has cheerful memories of being drilled how to correctly conjugate verbs and decline nouns.

"artificial" in this context because it suggests that speaking the language, or even the language itself, has an artificial feel to it. Strikingly, Ludwig Wittgenstein opposed Esperanto due to its alleged artificiality and occasionally had heated disputes about this topic with Rudolf Carnap, who was a proponent of Esperanto (cf. Löffler, 2004). Wittgenstein was not unique in his attitude towards Esperanto, but, like everyone who shares this assessment of the International Language ("Internacia Lingvo" was the original name for Esperanto before Zamenhof's alias "Dr. Esperanto" caught on as a name for the language), he was very poorly informed about the language. Hardly anybody who has actually learned the language claims that Esperanto has an artificial feel to it. In any case, Wells (2009, p. 375) at least makes clear that "[...] the epithet 'artificial' is arguably no longer applicable" to Esperanto if it ever was applicable in the first place. Be that as it may, we can put on record that "[...] Esperanto generally satisfies the criteria for recognition as a form of *natural* language." (Wells, 2009, p. 375; emphasis added)

I take it as settled that Esperanto is in fact a language, not merely some kind of language-like communication system, or whatever flimsy differentiations some occasionally try to introduce in order to deny certain conlangs their entitlement to a place among the "real" languages. "It must be emphasized that Esperanto is a real language, both spoken and written, successfully used as a means of communication between people who have no other common language." (Wells, 2009, p. 375) But what about the claim that Esperanto also has a distinctive culture?

5.2.4 The Culture of Esperanto

Being a fluent speaker of Esperanto myself and having participated in activities of the Esperanto language community during a couple of years,[12] I think that Arika Okrent (2009) provides a fair characterization of this language community. She takes an outsider's perspective on the community,

[12] It feels very roundabout to say that I "participated in activities of the Esperanto language community." It would be much more accurate to simply state that I participated Esperanto *culture* for quite a while, but I do not wish to beg the question whether there actually is an Esperanto culture at this point, of course.

and I can sympathize with her lively and credible depiction of some meetings and gatherings she witnessed. Although Okrent (2009) seems willing to leave the question open whether Esperanto really sparked its own culture, she eventually speaks about the culture of Esperanto without reservations.[13]

Okrent (2009) does so for good reason, and she cites some of the perennial statements which always come up to make a case for a distinctive Esperanto culture, in addition to an Esperanto language. There is plenty of literature in Esperanto, not only translated works, but also a fair amount of texts originally written in Esperanto, covering the whole span from poetry and fiction to technical manuals. There is a rich variety of music in Esperanto (originally composed by Esperanto speakers, with lyrics originally written in Esperanto, and performed in Esperanto, frequently during dedicated Esperanto music festivals for an Esperanto audience), ranging from folk-like via electronic (dance) music to reggae-style rap, with the repertoire's gravitational center inspired by classic rock. I draw on my own acquaintance with the scene in this description, and, since I am neither an expert in music nor its theory, I might have misclassified some examples I had in mind while writing this characterization. Nonetheless, my incompetence in this regard does not undermine the claim to cultural richness, shown by the Esperanto language community.

Moreover, there are several native speakers of Esperanto, speakers who learned Esperanto as their L1, i.e., their first language. Although, to the best of my knowledge, there does not exist any monolingual speaker of Esperanto (cf. Bergen, 2001, p. 576), it is not too uncommon to meet native speakers in the Esperanto community. Bergen (2001, p. 576) provides the number of "[. . .] 350 or so documented cases of Esperanto taught to children as their L1" in the late 20$^{\text{th}}$ century. This number seems to have been taken from Corsetti (1996), so the data I quote are not the latest. Given that the rapid spread of internet access pushed the popularity of Esperanto and made the number of its speakers significantly increase, current numbers of native Esperanto speakers can plausibly be estimated to be notably higher nowadays.

[13] Other scholars, e.g., Corsetti (1996, pp. 264 and 270), talk about "the culture of the Esperanto community" and "Esperanto culture" without any qualms.

Also Corsetti (1996) readily admits that the numbers were probably higher already in the mid-1990s, and he states that "[he] would not be surprised if the real number proved to be around one thousand, because every day new [Esperanto-speaking] families appear from nowhere." (Corsetti, 1996, p. 265) Be that as it may, the crucial question is not, I suppose, how many native speakers of Esperanto there exactly are. The more relevant aspect is that they exist, which they do. I am personally acquainted with some of them. Wells (2009, p. 375) even states that

> [t]here is no other case [than Esperanto] in linguistic history of something that started as an intellectual scheme, a project on paper, being transformed into a language with native speakers of the second and, indeed, the third generation.

I take all this unmistakable evidence for cultural life – the presence of original literature, music, radio (and even a few TV) shows,[14] and native speakers – as a given and will rather try to convince the reader of an actually existing and living Esperanto culture by drawing on some lesser-known linguistic pieces of evidence.

5.2.5 Esperanto Language and Culture

Leaving aside the elusive question whether language development follows cultural development or rather the other way round, there are several instances where linguistic developments and cultural peculiarities strikingly align in Esperanto. It is probably unsurprising that a community which is bound together by its language in a very explicit manner, as it is the case for speakers of Esperanto, also develops a distinctive vocabulary to serve its needs. Although the community is quite language-centered, it is far from hostile to languages other than Esperanto. In order to better understand the attitude most speakers of Esperanto hold to other languages, as well as typical self-conceptions of Esperanto speakers, we need to highlight the general situation

[14] Corsetti (1996, p. 270) at least mentions "[...] creative traditions in literature, theatre and music."

the Esperanto language community faces.[15] Speaking Esperanto constitutes a deliberate choice for the average speaker.[16] This choice includes the decision to live in a freely chosen diaspora (cf. Wells, 2009, pp. 375 and 376 f) since Esperanto is (deliberately) not linked with any country or place. Still, there is the expression "Esperantujo," which might be translated as "land of Esperanto".[17] The term, of course, does not refer to any country but rather to the sum of all places and institutions where Esperanto is used. This means, wherever people come together and use Esperanto to communicate, there is Esperantujo. The classical "place" to refer to with this word is (often international) conventions where speakers and speakers-to-be of the language gather.

5.2.5.1 Some Words on Code of Conduct

Whether any language but Esperanto is spoken during such meetings will, of course, depend on the proficiency of individual participants. Despite living somewhat in the shadows, the Esperanto community is quite open and welcoming, so that nobody who could not get through a conversation entirely in Esperanto will be left behind. Still, if it is not to integrate someone, speaking a language other than Esperanto in Esperantujo will usually be

[15] Some interesting remarks about the situation and self-image of Esperanto speakers – with a special focus on Esperanto-speaking families – can also be found in Corsetti (1996).

[16] Since native speakers are the minority, I leave them aside here.

[17] Translating "Esperantujo" as "land of Esperanto" certainly counts as correct, but there is a – I think – quite interesting trade-off between accurate and precise translation to be noted here: An accurate translation, I would say, captures the intended meaning when the expression is used. This is the case when we translate "Esperantujo" as "land or country of Esperanto." That said, a precise translation, which provides the *literal* meaning of a term, would come out less specific than "land or country of Esperanto." The suffix "-uj-" has a generic meaning of container (for something). By using this suffix you can form, for example, "ashtray" (cindrujo) from the word for ash (cindro), or "trash can" (rubujo) from the word for garbage (rubo), or simply the word for container (ujo). A further example is the word "gufujo," discussed on page 125. This kind of trade-off between what I called accurate and precise meaning of a term is very typical of Esperanto, due to its agglutinative structure and the rich word formation system, noted in footnote 27 on page 126. Translating "Esperantujo" as "land of Esperanto" is, however, licensed by the somewhat old-fashioned but still common practice to derive names of countries from the names of their inhabitants by using "-uj-" as, e.g., in "Francujo" (France) from "franco" (Frenchman).

frowned upon. On the other hand, not every situation of not speaking Esperanto in Esperantujo is alike, and here we need some peculiarly nuanced differentiations, which are mirrored in colloquial Esperanto.

As I mentioned before, not speaking Esperanto in order to include someone of low or no proficiency at all will always be accepted. If there is a veteran speaker of Esperanto around who also speaks the native language of the novice or visitor, this will usually be the language of choice to explain whatever needs to be expounded. However, it is a very different situation if two (or more) able speakers of Esperanto are caught in the act of chit-chatting in their native language without an exculpation like the integrative one mentioned before. This is called *"krokodili"* in Esperanto – roughly meaning, to speak a national language (in an international or Esperanto context) for no good reason (cf. Krause, 1999, p. 400). You can imagine that this comes close to committing a cardinal sin in an Esperanto meeting, and the perpetrator can only make amends by immediately explaining to the prosecutor what the uncovered conversation was about in plain Esperanto. (Do not even try to explain or even justify why you krokodilis[18] because there simply is no excuse for doing so.)[19] So, krokodili is evidently a bad thing to do. Every speaker of Esperanto knows that, and there is hardly any disagreement on the question whether a particular instance is a krokodilaĵo – i.e., a concrete case of committing what is called "krokodili" – although it is not trivial to spell out the exact criterion of what counts as a krokodilaĵo and what does not.

Remember that speaking a national language when talking to someone who does not understand Esperanto is never regarded as krokodilado,[20] and other kinds of circumstances might also make something not a krokodilaĵo though it would have been one in a slightly different context.[21] So, there is

[18] This is the past tense of the infinitive form "krokodili."

[19] The community is generally quite relaxed, so all of this is a bit of an overstatement. That said, given that nothing gets eaten as hot as it gets cooked (as a German proverb says, meaning that rules tend to be stated more strictly than they get applied), I would say that the depiction above adequately points out the normative demand even though real-life situations will turn out more lenient.

[20] The nominalization of the verb "krokodili."

[21] There certainly are some unclear cases, such as a situation where native Esperanto

a quite complex code of conduct only related to the question of when it is appropriate to speak which language in Esperantujo. However, we are not done yet with the dos and don'ts regarding when to speak which language in an Esperanto context. There is not only krokodili, which is always a bad thing to do, but there also are the expressions *"aligatori"* and *"kajmani,"* where matters become a little more complicated.[22]

These latter terms respectively name situations where speakers converse in a language (other than Esperanto) which is a second language for every speaker involved, and situations where the language spoken (other than Esperanto) is at least not the native language for every participant of the conversation. Using second languages other than Esperanto is often even encouraged in Esperantujo in the name of linguistic diversity and pluralism. Consequentially, dedicated spaces and times to speak other (second) languages are often provided during Esperanto meetings – the so-called aligatorejoj, i.e., places to aligatori[23] – to promote opportunities for partici-

speakers speak their mother tongue – I mean Esperanto, of course – in the presence of people who do not understand their language. Is that a krokodilaĵo? Maybe. The answer to that question will certainly depend on whether you frame your concept of krokodili along the lines of speaking a *national* language which happens to be your first language, or along the lines of speaking your *native* language. As with most expressions of the vernacular, there is no precise definition available to settle the issue for "krokodili." It is an expression which serves concrete but sometimes fuzzy everyday needs of a certain community, not a technical term designed to communicate a clear-cut concept a scientist would desire.

[22] I honestly do not know where this fixation on large semi-aquatic reptiles comes from. To the best of my knowledge, no one really knows that. But the trend cannot be denied. Apart from the expressions already mentioned, there also exist some less common examples: "reptiliumi" is the umbrella term to cover all of these peculiar expressions. Moreover, there also is "lacerti," which means the very special occupation of speaking an invented language other than Esperanto in Esperantujo. Finally we have "gaviali" to refer to situations where Esperanto is spoken instead of a different, more adequate language for the context in question. For an overview, see the Wikipedia entries on "krokodili" (Wikipedia contributors, 2019a) and "reptiliumi" (Wikipedia contributors, 2019b) on Vikipedio, the Esperanto version of Wikipedia on `http://eo.wikipedia.org`. Unfortunately, none of these entries is translated to any other language, but Google translator has a good enough command of Esperanto to help out – although the meaning of the English translation is sometimes slightly distorted, so that caution is required. For a quick overview of the relevant expressions in English, which attributes slightly different meanings to the expressions discussed in the main text, see Schor (2007).

[23] There are occasional controversies among speakers of Esperanto about the correct meanings of the expressions "aligatori" and "kajmani" to the effect that some claim that

pants to practice their non-native language competencies. Such lounge-like spaces with a dedicated purpose are not uncommon in Esperantujo. Another widespread example is the *gufujo*,[24] which serves as a quiet place for those tired of partying in the late evening to enjoy the tranquility of a calm conversation with a decent cup of tea.

5.2.5.2 Some Evidence for the Existence of a Living Tradition

The expressions discussed are clearly motivated by cultural needs, and most of them are, as far as I know, unique to Esperanto. This should be a good reason to infer that there also exists a unique Esperanto culture. Another term which could hardly exist without an underlying culture (and history) is the word *"kabei,"* which shares some similarities with the English expression "to gerrymander" in its origin, although not in its meaning. "Kabei" is the verb form[25] of the name of a once famous figure in the early Esperanto movement with the pen name "Kabe." Kabe was a very influential and prominent figure within the Esperanto community during the early 20$^{\text{th}}$ century until he suddenly turned his back on the language and completely disappeared from the movement without providing any reason for his decision. Correspondingly, the word "kabei" is nowadays often used with the meaning "to surprisingly and completely desert something" but also still used with the more precise meaning "to surprisingly and completely desert the Esperanto movement after having been an active part of it."[26]

Finally, allow me to add one more, rather anecdotal, piece of evidence

what you are actually supposed to do in an aligatorejo (i.e., a place to speak neither Esperanto, nor your mother tongue) is actually kajmani. Still, in general, the more consistent use of "aligatori," for what an aligatorejo is meant to be used for, prevails.

[24] The name is based on the Esperanto word for an eagle-owl ("gufo") to evoke associations of a nocturnal person, a night owl, reinforced by the stereotype of an owl as a rather silent creature. For an explanation of the suffix "-uj-" in "guf*uj*o," see footnote 17 on page 122.

[25] You can basically transform every word into any word class in Esperanto, given that the transformation makes sense.

[26] The Wikipedia article for the term "kabei" (cf. Wikipedia contributors, 2020a) is, again, not available in English, unfortunately. There is, however, a short but still informative article about Kabe (the person) in English, which provides "to fervently and successfully participate in Esperanto, then suddenly and silently drop out" (Wikipedia contributors, 2020b) as the meaning of "kabei."

for the existence of Esperanto culture. My entry into the Esperanto community was somewhat unusual. Since the language has a very straightforward structure and an extremely powerful apparatus for word formation, it allows for a quick entry into actual conversation by the newcomer student without arduous drilling in vocabulary and grammar. Consequentially, many people just tend to drop into an Esperanto meeting with rudimentary language skills, and they usually do just fine by picking up what they lack from direct conversation with other Esperanto speakers. So, a good formula for successfully learning Esperanto is: Just acquire the fundamental grammatical rules (there are just 16 of them), memorize some basic vocabulary and the most important word formation syllables, and you are good to go.[27]

For me, however, the story of becoming a member of the Esperanto community was quite different. Between my initial attempt to learn the language and my first Esperanto meeting, several years passed which were filled with a decent amount of reading in Esperanto. Due to this fact my vocabulary was packed with many rather literary expressions, which make a good appearance in the *belles lettres* but hardly ever come up in everyday conversation. I therefore initially encountered some rather unusual communicative barriers in Esperanto because I tended to use rare expressions which even highly experienced speakers of Esperanto hardly knew. I would, for example, use the term "krepusko" (designating the twilight during dusk and dawn) instead of the proper, common expressions "sunsubiro" and "sunleviĝo" to talk about sunset and sunrise. It takes a fair amount of reading to ever encounter a term like "krepusko," so my initial entry into the Esperanto community suffered some odd language barriers.

However, the point is that facts about prevalence of expressions in different contexts – e.g., everyday conversations in contrast to literary language – is a pragmatic factor which arguably supervenes on cultural aspects. I decently mastered the language when I made my first encounters with other speakers of Esperanto, but I still had to learn how people actually talk, *viz.*, I

[27] A textbook with the core word pool for Esperanto comprises approximately 600 entries, claiming that these will provide you (due to the powerful word formation system of Esperanto) with a *de facto* lexicon of between 3000 and 4000 words, which will allow a speaker to understand 98 % of any normal text (cf. Mayer, 1992, p. 5).

had to catch up on a substantial amount of contemporary Esperanto culture. Given that much of my early reading in Esperanto consisted of translated 19th century novels,[28] I suspect that my situation was in some sense similar to that of a person who learned English in the later 20th century by acutely studying John Locke's writings.[29] You speak the language, but your way of talking will strike your contemporaries as sorely peculiar.[30]

5.2.5.3 What to Make of All the Evidence for Esperanto Culture?

The previous illustrations of Esperanto culture were, of course, not only listed to merely amuse the reader. They are part of an argument against the claim that being socioculturally embedded is a necessary feature of language. The compilation of facts about the Esperanto language community which filled the previous pages is meant to demonstrate that a dedicated Esperanto culture does indeed exist. This culture is different from any other culture although it plausibly emerged from the melting pot every international community – just like the Esperanto language community – represents. It is probably impossible to state an exact date when this culture came about, but, every plausible date which is even worth considering is clearly later than the official birthday of the Internacia Lingvo (i.e., Esperanto – see the explanation on page 119) in 1887 when its foundational document was published. The crucial point is that the language in this case clearly predates its culture. Therefore, the language existed before its culture did. But how could that be if a language is necessarily embedded in its culture? The solution to this "riddle" is very simple, I claim: It is plainly not the case that a language is

[28] Although this undeniably brings in a temporal aspect, we are not dealing with linguistic evolution or change. The matter in question is rather an issue of learning a different language register.

[29] This is not a made-up example but a true story. I have heard a first-person report of someone's experiences with exactly this background. The person in question will probably know who is meant, and I want to take this occasion to express my sincerest and deep gratitude to this human for support and exemplary effect which really made a difference in my life – and, of course, also for sharing the anecdote in question with me.

[30] Along with these rather anecdotal pieces of evidence, we may also add that certain frequent deviations from correct standard Esperanto in spoken everyday Esperanto "[...] can be interpreted as an indication that the language is truly socially embedded ('living')." (Wells, 2009, p. 376)

necessarily or essentially socioculturally embedded, at least not in its own culture.

In order to argue for the existence of a distinct Esperanto culture, I extensively drew on several peculiarities of the Esperanto community to show that it is governed by a quite distinctive code of conduct which arguably amounts to a self-standing culture. Especially while discussing terms like "krokodili" and its cognates, I explained what they mean by exposing a certain cultural background and norms of behavior connected with these expressions. A certain suspicion might easily arise in many readers at this point: Did I not undermine my own claim that language and its culture are not inextricably intertwined by this very argumentative strategy to explain what certain expressions mean by exposing their cultural background? I do not think that this is the case, and I will explain my reasons for thinking so by presenting an analogy.

Just as I probably cannot properly understand a term like "krokodili" without some basic insights into the respective culture, I also cannot properly understand an expression like "particle accelerator" without a fair amount of knowledge about physics. Obviously, it would be absurd to claim that you therefore cannot understand a language without studying physics. Still, if this is so, why should we think that you cannot understand a language without studying its culture? An explanation of what it means to krokodili is best achieved by certain cultural clarifications just as an explanation of what a particle accelerator is requires knowledge of physics. Likewise, explaining what a gross domestic product (GDP) is demands at least rudimentary familiarity with economics. That said, if the latter do not warrant the conclusion that you cannot understand or study a language without also investigating physics and economy, respectively, then the former also cannot warrant the claim that a certain culture needs to be studied in order to understand a language.

5.3 Back to the Claim of the Necessary Sociocultural Embeddedness of Language

It should be clear that the preceding considerations prove that the claim that language cannot be separated from culture – introduced as *the claim of the necessary sociocultural embeddedness of language* on page 92 – cannot withstand scrutiny in this crude form. The claim is in fact so vague that it is hard to imagine a cogent argument which could refute it. A precisification of the claim that there cannot be language without culture is therefore required. The two most obvious ways of amending the claim in question, I think, are plausibly the following two. Either one might claim that

(1) there cannot be a *national* language without *its* respective culture,

or one might claim that

(2) there cannot be a language without being grounded in some culture, not necessarily its own.

The first modification only covers a specific class of languages and is therefore not a claim about language *per se*. The restriction to national languages is one plausible qualification among possible other options. However, the class of national languages is, linguistically speaking, an entirely arbitrary category. Even if claim (1) could be defended, it can hardly extend our understanding of the notion of language.

The situation would be different if we could claim, e.g., that no language could be a *native* language without there being a corresponding culture – under the assumption that the property of being a possible native language is not as contingent as the property of being a possible national language. However, this claim is probably not true. We can certainly imagine a child being raised in Esperanto in the late 19$^{\text{th}}$ century before any Esperanto culture plausibly emerged. Also, an American linguist actually raised his son in an invented language, speaking exclusively Klingon to him during the first three years of his life (cf. Bannow, 2009; Hiskey, 2012).[31] If the child had not

[31] The original case mentioned here apparently even found an imitator; cf. Coles (2018).

lost interest, we might now have had a native Klingon speaker. It does not seem too hard to imagine this scenario, but it would not change anything about the ontological status of the Klingon culture we know from television. There would still be the fictional culture known as Klingon though it would not transform this fictional culture into an actual culture.

The second precisification (2) is meant to cover all languages and can therefore plausibly be seen as a claim about language in general, but it probably amounts to a truism. I am willing to grant that language – and therefore also any particular language – is a cultural good, and it is certainly unsurprising that you cannot have a cultural good without there also being a culture. A language, being a cultural good, comes into existence when it is invented by cultural creatures, either individually as in the case of many invented languages or by an implicit and communal process of a community. Claim (2), I suppose, is not what most proponents of the inextricable entanglement of language and culture have in mind. Rather, I suspect, the stronger claim (1) – but probably without the restricted application to national languages only – is what many people find themselves drawn towards. Let us call this claim

(1') Every language needs to be embedded in its *own* culture.

This is the claim which, I think, looms in the background of a large proportion of the linguistic relativity discussion; and this is the claim I take to be conclusively refuted by the considerations presented here.

5.4 Conclusion

To recapitulate, I offered two ways to argue for the claim that language is not necessarily socioculturally embedded by offering counterexamples, i.e., languages which are not socioculturally embedded (in the right way). The first line of argument claims that the Klingon language, as an example of languages from fiction, represents a counterexample to the claim to be refuted. Since there is no Klingon culture (although there is a fictional Klingon culture), the very existence of the Klingon language (and other languages from

fiction) refutes the presumption that sociocultural embeddedness is a necessary feature of language. Klingon has no culture, which proves that a language can perfectly well exist without a corresponding culture.

The second line of argument builds on the example of Esperanto. While Esperanto arguably has its own culture, the language was clearly there before its culture came about. Consequently, the second line of argument proves exactly the same point: A language can perfectly exist without a corresponding culture. That the language was created in a cultural context is irrelevant as I do not wish to argue against (2). Klingon arose in a context of the American culture since its creator is part of the American culture. Esperanto was invented by a Polish Jew who grew up in the Russian Empire. So, there are quite a few candidates available for the position of being the culture which sparked the most successful constructed language. Be that as it may, neither the Jewish culture, nor the Polish culture, nor the Russian culture, nor any other culture with a tradition of much more than 130 years is *the* culture of Esperanto. We therefore have two counterexamples to claim (1') and by the same token a solid demonstration that it is not true that a language needs to be embedded in its own culture. Independently of this claim, it might be a quite plausible assumption that the continuous use of a particular language as a means of actual human communication will inevitably give rise to a corresponding culture, but this does not help to save the claim that language needs to be embedded in its very own sociocultural context from refutation.

Part III

Can We Say Everything We Think and Think Everything We Say?

An Investigation Into the Principle of Expressibility

Chapter 6

Introduction

A first clarification of the question which governs this third part of the present investigation – namely whether we can say everything we think and think everything we say – is needed right away in order to prevent a possible and fundamental misunderstanding: The word "can," as it is used here, has no relation to moral considerations whatsoever. The question could mistakenly be read as asking whether it is morally permissible or socially acceptable to say everything we think. This is not the intended meaning. Rather, "can" needs to be understood as relating to (metaphysical) possibility or feasibility. Is it possible, at all, to say whatever might come to one's mind? And is it possible, in general, to understand everything that someone might utter – including the question whether we ourselves can understand everything we might say? This is the guiding question of the present part while questions regarding social or moral appropriateness will be entirely ignored in what follows.

We will be concerned with an attempt to elucidate the relation between language and mind by considering the relevant domains as sets in order to investigate their relationship in terms of set theory. This constitutes a different and independent approach to discussing the topic at hand – i.e., the relation between mind and language – since the crucial distinction between thinking and reasoning (established in section 1.3) can be neglected in what follows. What is under consideration here is the entire domain of thought

(i.e., whatever may enter a speaker's stream of consciousness) and its relation to language. The discussion of this view on the matter shall be guided by John Searle's *Principle of Expressibility* (cf. Searle, 2011, pp. 19-21). Before this principle can be appropriately clarified and evaluated, the background for the discussion in the aforementioned set theoretic terms should be laid out.

6.1 Setting Up the Debate

In order to discuss the relation between language and thought/mind in this way, we can talk about *what can be said* or expressed – which comprises the domain of language in a set theoretic perspective – on the one hand, and contrast it with *what can be meant* or thought – the set theoretic analogue for the domain of the mind – on the other hand.

6.1.1 Logically Possible Configurations

After distinguishing what can be said from what can be meant in this way, we can list the logically possible configurations these two sets (the "sayable" S and the "meanable" M) can stand in with each other:

1. M and S coincide or are equivalent: $M \equiv S$
   This means that everything which can be thought can be said, and everything that can be said can be thought. Conversely, this also means that there is nothing that can be thought but not said and nothing which can be said but not thought.
2. S is a proper subset of M: $S \subset M$
   This means that there are things which can be thought but not (adequately) expressed, but everything which can be said can also be thought.
3. M is a proper subset of S: $M \subset S$
   This means that there are things which can be said, but not (honestly) meant, but everything that can be said can also be thought.

4. M and S are disjoint sets: $M \cap S = \emptyset$
   This means that nothing we can think can be (properly) expressed and nothing which can be said can be (honestly) meant in this way.

5. M and S overlap or intersect: $M \cap S \neq \emptyset$
   This means that some things can be expressed exactly as they are meant and be meant exactly as they are said, but some thoughts cannot be (adequately) expressed, and some things we can say cannot be (honestly) meant in this way.

These five options exhaust all possible relations in which M and S can stand to each other. All possible configurations are additionally depicted by means of Venn diagrams in figure 6.1 on page 140 for ease of exposition.

6.1.2 Sentence Meaning (S) and Speaker's Meaning (M)

It is common practice to distinguish between sentence meaning and utterance meaning in philosophy of language (cf. Searle, 1979c, p. 143). While sentence meaning (being the literal meaning of a linguistic expression) can, for the moment,[1] be straightforwardly identified with what can fall under S, the identification of utterance meaning with what falls under M is more problematic. Since M is supposed to comprise everything which can be thought independently of any attempt to be ever expressed, the term "utterance meaning" is unfortunate insofar as it might suggest that only meanings which are at least attempted to be expressed fall under this label.

So, "utterance meaning" plausibly only covers meanings a speaker intends to convey with a certain utterance. The expression "speaker's meaning," which is sometimes used interchangeably with "utterance meaning," fits better for our purpose. The expression "speaker's meaning" can cover any content a (potential) speaker has in mind even when she does not make an utterance and is (currently) not making any communicative attempts. Since

[1] Ultimately, sentence meaning and set S will not be associated so closely. *Utterance meaning* is the correct category to be identified with what can fall under S. However, these intricacies will be ignored for the moment. See the final paragraph of this section (6.1.2) on the next page.

only the possible mental contents of linguistic creatures will receive consideration here, the term "speaker's meaning" does not introduce any additional restriction to the scope of M.

John Searle (1979a, p. x) also contrasts "[...] literal sentence meaning and intended speaker's utterance meaning;" or even "[...] speaker meaning and literal sentence meaning [...]" (Searle, 1979a, p. xi) in order to make "[...] the general distinction used throughout this book between the meaning of the expressions that a speaker utters and his intended meaning [...]" (Searle, 1979a, p. xii). This is also the contrast I wish to make between (literal) sentence meaning and speaker's meaning. I will, however, not follow the common practice to identify speaker's meaning and utterance meaning, nor will I use the terms "speaker's meaning" and "utterance meaning" interchangeably. I think that utterance meaning should be distinguished from sentence meaning, as well as from speaker's meaning. However, since this distinction is not relevant at the moment, I will not introduce it now. For the moment, only the distinction between speaker's meaning (i.e., the meaning/content intended by a speaker) and sentence meaning (i.e., the literal meaning of the words uttered) is relevant. I will introduce utterance meaning only when it is needed, at a later stage of the discussion, in section 9.1.3 on pp. 230 ff.

6.1.3 Background Considerations

It shall be assumed, for the sake of the present investigation, that what will be sorted into the sets S and M are contents: namely the contents of what can be said and the contents which can be thought/meant respectively. The contents of what cannot be said and the contents which cannot be thought or meant need to find their respective spots as well. The content of what cannot be said will find its way into M, and the contents which cannot be thought will find their way into S. There will be no content left outside of both sets, *viz.*, nothing can be neither thought nor said.

This is not meant to be a substantive claim about what contents there are by denying the existence of contents which can neither be thought nor

expressed. There might be contents which cannot, for example, even be grasped by infinite minds. Also it might, for whatever reason, be impossible to express such contents, even for an infinite mind. I actually doubt that such contents (which neither fit into M nor into S) could possibly exist. But there is no need to defend this view at this point. Since the present investigation is merely concerned with the relation between what *can* be thought and what *can* be expressed, contents which can neither be thought nor expressed will simply be ignored in what follows – regardless of whether there are such contents.

It shall also be naïvely assumed that *we can sort* every relevant content into either or both of the sets S and M. Every possible content which is relevant for the present investigation therefore belongs to the members of at least one of the two sets under consideration. I will also, for the moment, presuppose that for every content it can be unproblematically decided to which set it belongs and that the sets in question, as well as their members, can be smoothly compared.[2]

6.1.4 Excluding First Options

Before we come to Searle's Principle of Expressibility, it should be noted that some of the five options mentioned on pp. 136 f, and depicted in figure 6.1 on the next page, represent merely logical possibilities.

6.1.4.1 Excluding Option 4

Option 4 (according to which what can be said and what can be thought do not have any common members) is a merely theoretical option. Since asserting that mind and language stand in the relation expressed by option 4 would amount to the claim that there is not even a single thought we can properly communicate, this possibility can be readily dismissed as absurd. If I honestly utter the sentence "The weather is beautiful today," then I mean exactly what I said. This mundane fact is sufficient to exclude option 4 from

[2] This topic will come up again on page 160, and whether this presumption is warranted will be discussed in chapter 10, on pp. 255 ff.

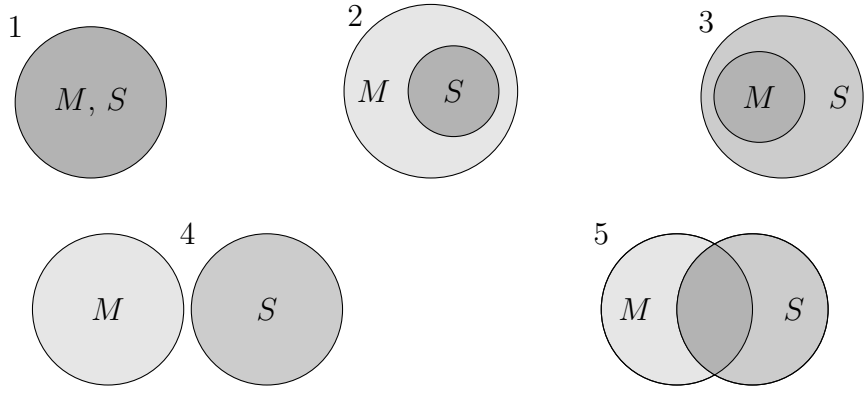


Figure 6.1: Possible relations between what can be thought or "meant" (M) and what can be expressed or said (S)

the list of possible relations between mind and language which are worthy of serious consideration.

6.1.4.2 Excluding Option 1

It also does not seem, on first glance, that hard to exclude option 1 (according to which what can be said and what can be thought perfectly coincide) from the list of plausible relations between what can be said and what can be thought. We can utter apparently nonsensical stuff, such as Chomsky's famous example "colorless green ideas sleep furiously" (Chomsky, 1965, p. 149),[3] which plausibly cannot be literally "meant" nor properly understood. Chomsky's sentence probably does not even correspond to anything we could consider as describing possibly comprehensible content. Therefore, option 1 can also be dismissed since the view that what can be said perfectly coincides with what can be meant implies that nothing can be said that cannot also be thought (or meant).

[3] Searle (1979b, p. 77) also makes use of this example. Everything we can say about Chomsky's example, discussed in the text, we can also say about "[...] Russell's example of a completely nonsensical sentence, 'Quadrilaterality drinks procrastination'" (Searle, 1979b, p. 92). I was not able to find the sentence Searle quotes in Russell's texts, but the nonsensical sentence "quadruplicity drinks procrastination" appears in Russell (1995, pp. 166, 170, 177, and 275). Russell (1995, p. 183) also talks "of quadruplicity killing procrastination." I guess that all of these, including Searle's variant, can serve equally well as examples for category mistakes.

On the other hand, Ofra Magidor (2009) provides strong arguments to the effect that category mistakes are not meaningless (as we would assume for a nonsensical utterance) but false.[4] Given that ideas do not sleep at all, we can plausibly conclude that it is also not the case that colorless green ideas sleep furiously. If we assume that the sentence "colorless green ideas sleep furiously" is false, then it also needs to be meaningful. If the sentence is meaningful, we can probably also think or mean what the sentence expresses. Otherwise, it would be hard to understand how we could possibly be fairly certain that what the sentence says is not the case. Notwithstanding the long tradition of the claim that category mistakes are meaningless, it is easy to sympathize with Magidor's view. We should therefore not rule out option 1 too quickly, and the possibility that the existence of category mistakes does not conflict with the claim that what can be said and what can be thought perfectly coincide should be seriously considered. Before we take a second look at option 1, however, we should quickly clear up two central notions in this discussion.

6.1.4.3 What Does It Mean to Say Something and to Mean Something?

Magidor's (2009; 2013) claim that category mistakes are actually meaningful aside, I think that we should extend the scope of (the set of) what can be said to nonsensical utterances in any case. This opinion, however, is in conflict with other possible understandings of what it means to *say* something. Frank Kannetzky (2001, p. 192), e.g., claims that "[u]ttering something that cannot be understood by anybody is to say nothing." In a different place, a couple of pages later, where he discusses an "ambiguity of 'to say something'" (Kannetzky, 2001, p. 201), Kannetzky adds that

> [...] 'saying something' can be understood as 'making it understandable'. One can mean only something that can, at least

[4] For more details on the discussion about category mistakes and the meaningfulness view Magidor (2009) endorses, see Magidor (2013) which is her book-length elaboration on the topic. I wish to thank Mark Textor for pointing me to Ofra Magidor's investigation and Frank Hofmann for interesting discussions in this context.

> in principle, be made intelligible to others. (Kannetzky, 2001, p. 201)

This understanding of what it means to say something would not be terribly useful for the investigation at hand. Consequentially, it is not this understanding of what it means to say something which governs selection of the members of the set of what can be said (S). Even more importantly, the restriction that "[o]ne can mean only something that can, at least in principle, be made intelligible to others" (Kannetzky, 2001, p. 201) would beg the question regarding the truth of the Principle of Expressibility, which we will take a first look at soon (namely on page 144).[5] So, also the selection of what is a member of the set of what can be meant (M) must not be governed by this understanding of what it means to say or mean something.

I recognize this possible use of the notion of saying something but deliberately choose a more inclusive understanding of what it means to say something and of what it means to mean something. Certain terminological choices could preclude any substantive discussion of how the sets M and S relate to each other. So, even thought the notions of what can be meant and said can be understood differently from how I use them, a very liberal understanding needs to be presupposed here. Otherwise, the discussion of the Principle of Expressibility in set theoretic terms might be trivialized by preliminary terminological decisions.

Although I opt for a liberal understanding of what it means to say something, we must not be too liberal in this regard. A minimal requirement that *only utterances with semantic features can count as saying something* needs to stay in place. This restriction shall exclude that uttering something like "banderwooky biggley booba toves," or whatever alphabetical strings may appear on your computer screen after you fell asleep on your keyboard,[6]

[5] This understanding of what it means to mean something would especially beg the question against the *strengthened* Principle of Expressibility, which will be introduced in section 6.2.4 and represents the primary concern of this investigation.

[6] Note that I ignore the highly unlikely event that your resting body accidentally activates the keys of your keyboard in a way which produces a perfectly grammatical and meaningful text. In this case, I suppose, it should be uncontroversial to say that what appears on your screen can not only be said but also meant and understood. The more likely

count as saying something (in English).

6.1.4.4 Proceeding With Option 1, and Excluding Option 2

The mere existence of category mistakes is at least not obviously sufficient to exclude option 1 since it is plausible to assume that whatever is meaningful can also be thought. Category mistakes might represent an exception to this general principle if we can find a cogent argument for the possibility that we can legitimately qualify sentences as being false even without the ability to think the expressed content.[7] Yet we have other options to refute the view that what can be said coincides perfectly with what can be thought (i.e., option 1). The reader who remains unconvinced might also consider utterances which do not even purport to express a thought that can be entertained. I can readily say something like, e.g., "coincide but and later not car dense." This utterance certainly counts as saying something in the sense specified in the previous section 6.1.4.3 since the utterance has semantic features due to the fact that every constituent of the utterance has semantic features. Still, can I think or mean what I said? It is far from obvious what could even count as (honestly) meaning what I said in such a case insofar as what I uttered is just an arbitrary list of words which do not interact to form a sentence which could express a conceivable content.

I can, of course, think of what the words "coincide," "but," "and," "later,"

result of your resting on the keyboard, however, will be a meaningless string of characters like this one: `...iOdfpjöÄ‘ÜFO?IK....` It seems safe to claim that this cannot be said, let alone meant or understood.

[7] It might be sufficient to know which proposition is expressed without being able to actually think or entertain said proposition in order to determine the truth value of a sentence. Or it might not even be necessary to know which proposition is expressed. I know, e.g., that the sentence "I am here now" is certainly true even if I do not know to whom "I" refers, nor to which place "here" refers, nor to which point in time "now" refers (cf. Kaplan, 1989, pp. 508 f; see also Salmon, 1986, p. 177, n. 1 & 2, and pp. 179 f, n. 19). I therefore arguably do not know which proposition this sentence expresses, but I nevertheless know that it is true. So, to say that we can know the truth value of a sentence without the need to (be able to) entertain the content expressed is certainly not out of question. This would allow us to say that category mistakes are meaningful, since false, while still leaving the option available that we can know all that without being able to think what a category mistakes expresses. Magidor's (2009; 2013) claim that category mistakes are meaningful is therefore not necessarily in conflict with the position that we cannot think or mean what category mistakes express.

"not," "car," and "dense" mean in turn – they have semantic features, after all – but I cannot think *that coincide but and later not car dense*. Even if we assume that I can really think that colorless green ideas sleep furiously, it is a considerably harder bullet to bite that I can also think that coincide but and later not car dense. What kind of situation could I even have in mind which can be described in these words? I presuppose here, of course, that the words in question have their usual meaning and take their place in common English. If the sequence of words under consideration was, e.g., part of a special code to communicate a covert message, the situation might be different. Under very specific circumstances, an utterance of "coincide but and later not car dense" might, for example, inform a spy that she was uncovered and needs to keep a low profile until she can be evacuated. Yet this is not what such an utterance means in "normal" English. In plain English, this specific sequence of words does not mean anything, and this is the situation we are interested in. Counterfactual situations where an utterance of just these words in just this order might mean something (different from what they actually mean) are irrelevant in this context and can therefore be neglected in the context of the question whether option 1 correctly depicts the relation between the sets M and S.

These considerations, by the same token, also exclude option 2. If we can say things which we cannot understand or think, then what can be said is not a proper subset of (or included in) what can be thought or meant. This leaves us only with options 3 and 5 as possibly tenable positions for how mind and language might in fact be related.

6.2 The Principle of Expressibility

Before we can decide whether option 3 or option 5 correctly depicts how mind and language are related, we should introduce and take a closer look at Searle's Principle of Expressibility, which was already mentioned on page 136 and on page 142.

Searle's Principle of Expressibility is "[t]he principle that whatever can

be meant can be said [...]" (Searle, 2011, p. 19).

This principle is compatible with options 1 and 3 whereas it contradicts all other listed options.

6.2.1 The Principle of Expressibility and the Sets M and S

Since option 1 (the view that what can be said and what can be meant perfectly overlap without residue) was already excluded, a proof of the Principle of Expressibility would leave us with option 3 (according to which the scope of what can be thought is contained in the scope of what can be said but not *vice versa*) as the true relation between mind and language. But option 5 (which holds that there are things which can be said but not thought, as well as things which can be thought but not said) is, at least *prima facie,* also a plausible option. If option 5 turned out to be true, this would prove the Principle of Expressibility wrong. Before any decision in this regard can be reached, however, the Principle of Expressibility, as it is endorsed by John Searle, needs to be scrutinized and stated in more detail.

6.2.2 A First Closer Look at the Principle of Expressibility

In order to precisely formulate the Principle of Expressibility, Searle (2011, p. 20) writes that

> [a]ny language provides us with a finite set of words and syntactical forms for saying what we mean, but where there is in a given language or in any language an upper bound on the expressible, where there are thoughts that cannot be expressed in a given language or in any language, it is a contingent fact and not a necessary truth.
>
> We might express this principle by saying that for any meaning X and any speaker S whenever S means (intends to convey,

> wishes to communicate in an utterance, etc.) X then it is possible that there is some expression E such that E is an exact expression of or formulation of X. Symbolically: $\forall S \forall X (S \text{ means } X \rightarrow \Diamond \exists E (E \text{ is an exact expression of } X))$.[8]

In contrast to the informal and preliminary statement of the Principle of Expressibility as saying "[...] that whatever can be meant can be said [...]" (Searle, 2011, p. 19), this more detailed statement provides some noteworthy explications. Especially, the term "exact" does not show up in the preliminary formulation but is introduced in the more detailed statement just quoted. So, even the shortened and informal version of the Principle of Expressibility should probably be amended to say: "Whatever can be meant can be said *exactly*." (Cf. Hackstette, 1982, p. 425) We will come back to the question of what "exact" means precisely on pp. 152 ff.

6.2.3 Qualifications of the Principle of Expressibility

Searle urges the need to qualify even the detailed version of the principle immediately after having formulated it, in order "[t]o avoid two sorts of misunderstandings [...]" (Searle, 2011, p. 20).

Searle's first qualification concerns pragmatic effects. He emphasizes

> [...] that the principle of expressibility does not imply that it is always possible to find or invent a form of expression that will

[8] Notation adapted. Searle's (2011, p. 20) original formulation reads as follows:

> Symbolically: $(S)(X)(S \text{ means } X \rightarrow P(\exists E)(E \text{ is an exact expression of } X))$.

The formula, as given in Searle (2011, p. 20), precedes the following note:

> This formulation involves an explicit use of quantifiers through a modal context; but since the kind of entity quantified over is 'intensional' anyway, the modal context does not seem to raise any special problems. (Searle, 2011, p. 20)

"S" is here, evidently, a variable which ranges over any speaker, and not, as before, the name of a set – namely the set of everything that can be said. Context should always be sufficient to prevent any possible confusion, so I will not take any measures to explicitly indicate which S is meant, the variable (ranging over any speaker) or the set (comprising everything that can be said).

> produce all the *effects in hearers* that one means to produce; for example, literary or poetic effects, emotions, beliefs, and so on. (Searle, 2011, p. 20; emphasis added)

Illustrated with a markedly blunt example, this means that even if it may in principle be impossible, in any language, to express a dull idea in a way that convinces and enthralls at least one hearer, this will not render the Principle of Expressibility false. In evaluating the Principle of Expressibility we need to distinguish *what* a speaker means to convey (i.e., the content or meaning which shall be expressed) from *how* the speaker wishes to convey what she means (i.e., the pragmatic or perlocutionary effects a speaker wishes to achieve; cf. Austin, 1962, pp. 101 ff). Being perceived as eloquent, spirited, convincing, etc. concerns *how* something is communicated. The Principle of Expressibility, in contrast, only applies to *what* is communicated. The principle claims that no content is in principle ineffable, and it claims – in addition – not only that every content can be somehow conveyed but that every content can be *exactly* expressed in language. We will return to this point later, but, before we come to it (on pp. 152 ff), the second misunderstanding Searle wants to preclude deserves some attention.

Searle's second qualification: Searle (2011, p. 20) proceeds by saying that

> [... s]econdly, the principle that whatever can be meant can be said does not imply that whatever can be said can be understood by others; for that would *exclude the possibility of a private language*, a language that it was logically impossible for anyone but the speaker to understand. Such languages may indeed be logically impossible, but I shall not attempt to decide that question in the course of the present investigation. (Searle, 2011, p. 20; emphasis added)

Even if excluding the possibility of a private language is not part of Searle's program, it is a part of the investigation at hand. So, this qualification can be readily dropped in the current context. Since I endorse *PLA* (see section 1.2.2

on page 10) and its conclusion, according to which a language which can in principle be understood by nobody but a single speaker is indeed impossible, there is no need to follow Searle's example and hedge at this point.

Although it might not be clear that a private language is *logically* impossible, as Searle puts it in the quote provided above, I take it to be a demonstrated fact that there cannot be a private language in the Wittgensteinian sense. I am, however, open to the possibility – in fact, I even favor the view – that it is not logical impossibility, but rather metaphysical or conceptual impossibility,[9] which precludes a private language. Since I do not judge the question of very high importance in which way exactly a private language is impossible in the present context, I will not pursue this issue in detail. But it should be noted that not everything which is metaphysically or conceptually impossible is also logically impossible. Searle's conjecture that a private language may be *logically* impossible might therefore be false, while a private language is nevertheless perfectly – namely metaphysically or conceptually – impossible.

6.2.4 Consequences of Waiving Searle's Second Qualification: The Strengthened Principle of Expressibility

That we can exclude the possibility of a private language means that the Principle of Expressibility under consideration here is significantly stronger than the principle Searle argues for – even if it might not be stronger than

[9] I leave the question open whether and how metaphysical modality should be demarcated from conceptual modality. Yet the expression "conceptual impossibility" should not be misunderstood as indicating that a private language might be merely epistemically impossible. It seems rather obvious to me that a private language is at least *prima facie* conceivable and therefore at least in some sense epistemically possible. Otherwise, it would be an immense mystery why the claim that a private language is not possible sparked such a massive amount of discussion (cf. Raleigh, 2019, p. 70). The sense in which a private language is demonstrably impossible has to do with the (conceptual) structure of reality in a robust realist sense, in contrast to an idealist or transcendental understanding, i.e., independently of (how reality is perceived by) the human mind. (See also the quick discussion of what I call "solid metaphysical realism" on page 195 and especially the more detailed discussion in section 8.3.2.3.)

the principle Searle considers, given that metaphysical or conceptual impossibility is weaker than logical impossibility. Still, since I count the possibility of a private language as being refuted, we should take the Principle of Expressibility to imply that whatever can be meant or thought can, at least in principle, also be expressed *and* understood by someone else but the speaker.

It is important to note that this strengthened version of the Principle of Expressibility is not necessarily in direct conflict with Searle's second qualification of the principle, quoted on page 147. There, Searle insists that the Principle of Expressibility "[...] does not imply that whatever can be *said* can be understood by others" (Searle, 2011, p. 20; emphasis added). This restriction needs to stay in place. Given that we can say things which nobody – not even the speaker herself – can understand, it is certainly not true that "[...] whatever can be *said* can be understood by others" (Searle, 2011, p. 20; emphasis added). This is proven by nonsensical utterances, such as those mentioned on pp. 143 f in order to exclude options 1 and 2 from the list of plausible relations between language and thought.

The strengthened version of the Principle of Expressibility I wish to endorse does not claim that "[...] whatever can be *said* can be understood by others" (Searle, 2011, p. 20; emphasis added) but that whatever can be meant or thought – and in turn consequently also expressed if the Principle of Expressibility is true – can be understood by others. Since we already settled that the scope of the expressible (S) is neither contained in nor coincides with the scope of the "thinkable" (M),[10] it makes a considerable difference whether we claim that everything that can be *said* can be understood by others or that whatever can be generally *meant* can also be understood by others. The latter claim is what I take to be an interesting and plausible extension of Searle's Principle of Expressibility while the former is almost trivially false. So, waiving Searle's second qualification (see p. 147) on the basis of *PLA* from Part I, the strengthened Principle of Expressibility is the

[10] That M and S do not coincide (option 1) can be shown with the same example (see pp. 143 f) which proves that S is not contained in M (option 2). However, since option 1 is not in conflict with the Principle of Expressibility, excluding option 1 is not a priority of the present investigation. For orientation, see the illustrations in figure 6.1 on page 140 and the list of possible relations between the sets M and S, preceding figure 6.1, on pp. 136 f.

crucial principle which is ultimately under consideration here.

The *strengthened* Principle of Expressibility claims that whatever can be thought can be expressed in such a way that it can, at least in principle, also be understood by others.

6.2.5 Wittgenstein and the (Strengthened) Principle of Expressibility

None of this must be confused with Ludwig Wittgenstein's dictum that "what can be said at all can be said clearly [...]" (Wittgenstein, 2001, p. 3) which he gives expression to in the preface of his *Tractatus logico-philospohicus*. Given that what can be said (S) plausibly exceeds what can be thought (M), Wittgenstein's claim might appear to have a wider application than the Principle of Expressibility. As we have just seen, the strengthened Principle of Expressibility claims that whatever can be meant can be expressed in a way such that it can be understood by others, *viz.*, I take it, that whatever can be *meant* can be said clearly. Wittgenstein, in contrast, claims that whatever can be *said* can be said clearly. Since there are things which can be said but not thought, the scope of Wittgenstein's claim might be wider than the scope of the Principle of Expressibility, at least if we concede that even nonsense which cannot be thought can be expressed clearly.

Wittgenstein's claim is still considerably weaker than the Principle of Expressibility [11] because Wittgenstein explicitly takes into account that there are things which cannot be expressed. This becomes evident when Wittgenstein virtually continues his dictum cited above – which sums up 'the whole sense' of the *Tractatus* (cf. Wittgenstein, 2001, p. 3) – with the final sentence of his *Tractatus logico-philosophicus*: "What we cannot speak about we must pass over in silence." (Wittgenstein, 2001, § 7, p. 89)[12]

[11] Navarro-Reyes (2009, p. 303, n. 2) also makes this point and provides a similar discussion of the contrast between Wittgenstein's dictum and the Principle of Expressibility.

[12] The ultimate paragraph of the *Tractatus* reads as follows in the German original: "Wovon man nicht sprechen kann, darüber muß man schweigen." (Wittgenstein, 2003a, § 7, p. 111) The relevant passage in the preface of the *Tractatus* is:

> The whole sense of the book might be summed up in the following words:

Wittgenstein would therefore not subscribe to the Principle of Expressibility since, according to this principle (or at least its strengthened version), there is nothing (i.e., no content) we need to pass over in silence. If the Principle of Expressibility is true, Wittgenstein's distinction between "[w]hat *can* be shown, [but] *cannot* be said" (Wittgenstein, 2001, § 4.1212, p. 31)[13] becomes obsolete, and even *the mystical* can be put into words,[14] according to the (strengthened) Principle of Expressibility.

> what can be said at all can be said clearly, and what we cannot talk about we must pass over in silence.
>
> Thus the aim of the book is to draw a limit to thought, or rather—not to thought, but to the expression of thoughts: for in order to be able to draw a limit to thought, we should have to find both sides of the limit thinkable (i.e. we should have to be able to think what cannot be thought).
>
> It will therefore only be in language that the limit can be drawn, and what lies on the other side of the limit will simply be nonsense. (Wittgenstein, 2001, pp. 3 f)

In the German original, it says:

> Man könnte den ganzen Sinn des Buches etwa in die Worte fassen: Was sich überhaupt sagen läßt, läßt sich klar sagen; und wovon man nicht reden kann, darüber muß man schweigen.
>
> Das Buch will also dem Denken eine Grenze ziehen, oder vielmehr – nicht dem Denken, sondern dem Ausdruck der Gedanken: Denn um dem Denken eine Grenze zu ziehen, müßten wir beide Seiten dieser Grenze denken können (wir müßten also denken können, was sich nicht denken läßt).
>
> Die Grenze wird also nur in der Sprache gezogen werden können und was jenseits der Grenze liegt, wird einfach Unsinn sein. (Wittgenstein, 2003a, p. 7)

The passage suggests that, after all, things are not as simple as I presented them. If everything on the far side of what can be said clearly is nonsense, and nonsense – in contrast to what has no sense, i.e., tautologies and contradictions – is part of what "[. . .] we must pass over in silence" (Wittgenstein, 2001, p. 3), then Wittgenstein's notion of what can be said is much more restricted than what I count as falling under S, i.e., the set of what can be said – be it meaningful or not.

[13] Wittgenstein (2003a, § 4.1212, p. 40) reads as follows in the German original: "Was gezeigt werden *kann*, *kann* nicht gesagt werden."

[14] "There are, indeed, things that cannot be put into words. They *make themselves manifest.* They are what is mystical." (Wittgenstein, 2001, § 6.522, p. 89)
The corresponding passage reads as follows in the German original: "Es gibt allerdings Unaussprechliches. Dies *zeigt* sich, es ist das Mystische." (Wittgenstein, 2003a, § 6.522, p. 111)

6.2.6 A Critique of the Principle of Expressibility

Although the Principle of Expressibility is often considered to be "[...] a principle seldom articulated but commonly presupposed in contemporary philosophy of language" (Binkley, 1979, p. 307; see also p. 324), the principle did not go undisputed even in the weaker sense defended by Searle. Recall that this weaker sense did not insist on the impossibility of a private language and its near corollary that whatever can be meant and, according to the Principle of Expressibility, thereby articulated can also be understood. Timothy Binkley (1979) presents an interesting critique of Searle's principle. He claims that the expression "exact," as it is introduced in Searle's more sharpened formulation of the Principle of Expressibility on page 145, is ambiguous. This ambiguity, in turn, taints Searle's considerations and leads to certain errors in his account. While Binkley's (1979) critique remains unconvincing overall (for reasons to be explained in the following sections 6.2.6.1 and 6.2.6.2, as well as in section 6.2.6.4), he nevertheless presents a couple of striking considerations along the way.

One of these considerations is Binkley's (1979, p. 308) attempt "[...] to distinguish two different conceptions of exactness, [...]" namely *accuracy* on the one hand, and *precision* on the other hand. Binkley (1979, p. 309) explains the basic idea behind his distinction thus: "Accuracy is a measure of the lack of error in a particular reading; precision is a measure of the capacity to discriminate. Each can be understood as a kind of 'exactness'." Further on, he adds: "Accuracy is a matter of right and wrong, while precision is a matter of more and less." (Binkley, 1979, p. 309)

The following lines are particularly illuminating with regards to how the distinction between accuracy and precision as two kinds of exactness comes to play a role in application to Searle's Principle of Expressibility since

> there is a difference between saying exactly (accurately) what you mean and saying what you mean exactly (precisely). If we seek accuracy, we want to know simply whether the language expresses what the speaker means. Thus, strictly speaking, an inaccuracy occurs when there is a difference between what a speaker means

> [i.e., speaker's meaning] and what his utterance means [i.e., sentence or utterance meaning]. The paradigmatic case of linguistic inaccuracy is where the speaker makes a mistake in his use of language, usually characterized by the speaker's desire to change his expression when he discovers the error. The error could be one of many types: inadvertently saying the wrong word or phrase, using a word without fully understanding its meaning, making grammatical mistakes, etc. The precision of an expression, on the other hand, is not a matter of *whether* it expresses the speaker's meaning, but rather of *how* refined an expression it is. [...] For now it should simply be pointed out that precision will not have to do with whether an utterance says (accurately) what a speaker means, but rather with the degree of refinement in the expression he uses. Consequently, discerning precision will usually involve comparing two or more *expressions*, and will not involve comparing an expression with a speaker's meaning, as we do when investigating accuracy. The precision of a linguistic expression does not bear directly on the question of whether it expresses the speaker's meaning (accurately). (Binkley, 1979, pp. 310 f)

This somewhat longer quote nicely illustrates the difference between accuracy and precision Binkley (1979) wants to draw our attention to: To judge whether an expression is *accurate*, we need to compare the meaning of the expression with what the speaker means to say. By making this comparison, we can decide whether an utterance accurately expresses what a speaker wishes to say. Judging how *precise* an expression is, in contrast, does not require any reference to what a speaker means at all. In order to evaluate the precision of an expression, we compare it with other expressions to see which one is more precise. So, precision is a purely linguistic attribute and exclusively relates expressions to arrange them along a scale from more to less precise or refined. Accuracy, in contrast, goes beyond the domain of expressions since, in order to evaluate the accuracy of an expression, the meaning of an expression needs to be compared with something that might not be the meaning of an expression: namely the potentially unexpressed meaning intended by a

speaker. Note that therefore Binkley's distinction is not merely a distinction between binary and scalar exactness, so to speak.[15] The crucial difference lies in what is compared, not in the fact whether the comparison yields a yes-or-no answer or a more-or-less answer.

6.2.6.1 Binkley's Critique

Binkley (1979) builds his critique of Searle's Principle of Expressibility on the claim that it presupposes something that is not possible: comparing the meaning of an expression with the meaning a speaker intends to express. This is an impossible task, according to Binkley, because an unexpressed meaning cannot be compared to anything, or at least "[...] linguistic expressions cannot be compared with unexpressed meanings." (Binkley, 1979, p. 313) In order to have a meaning which can be compared with another meaning, the meaning needs to be expressed first. An unexpressed meaning, we might say, is too elusive an entity to allow for any comparison at all. In Binkley's (1979, pp. 313 f) words:

> The problem inherent in trying to discuss the accuracy of linguistic expressions is this: How do we tell what a "speaker's meaning" is independently of his expression of it? Unexpressed meanings are rather inscrutable entities. [...] We compare meanings by comparing expressions of meaning.

The critique that unexpressed meanings are a theoretically dubious kind of entity is a serious objection against the approach followed here. A frequent criticism against merely intended but potentially inexpressible meanings is that no identity conditions can be stated for such alleged meanings. Building on Willard Van Orman Quine's dictum, whereupon there is "No entity without identity" (Quine, 1969, p. 23), some scholars tend to reject the very idea of unexpressed meaning or content altogether. If this rejection of unexpressed meanings as theoretically eligible entities is warranted, the contrast between

[15] Thanks to Deven Burks for this way of framing a possible misunderstanding of Binkley's distinction.

what can be said and what can be thought (without being expressed) is in danger.

If we cannot separate the sets S and M for lack of a clear individuation procedure for the members of set M, then any attempt to elucidate the relation between language and thought by specifying the set theoretic relation between M and S might be doomed. If the very existence of the members of a set is questioned, then the set M is unfit to serve a serious investigation. This criticism needs to be considered, but, before we inspect in more detail the issue whether unexpressed meanings need to be expelled from any theory about the relation between mind and language, I wish to take a closer look at the question of how damaging Binkley's critique is for *Searle's application* of the Principle of Expressibility (which is different from the use I make of Searle's principle here in this investigation).

6.2.6.2 Saving Searle

When taken as an objection to the Principle of Expressibility as Searle uses it in his (2011) *Speech Acts*, Binkley's (1979) criticism can hardly be successful. I strongly tend to agree with Karl Hackstette (1982), who claims that, when Searle introduces the term "exact" into his formulation of the Principle of Expressibility, he neither means "accurate" nor (exactly) "precise,"[16] nor does he confuse the two. What Searle had in mind by "an exact expression" needs to be understood as an *explicit* expression. What "explicit" means in this context becomes clear when we take Searle's theory of speech acts into account.

I can issue the very same speech act either by saying

(I) I will come to your party tomorrow.

or by saying

(E) I (hereby) promise you to come to your party tomorrow.

[16] The reason for the parenthetical caveat at this point – namely why I say that Searle does not *exactly* mean "precise" – will be made fully explicit on pp. 159 f.

in an appropriate situation. The difference between (I) and (E) is neither the (type of) speech act issued, nor its content or its illocutionary force (cf. Austin, 1962, pp. 99 ff), but merely the degree of explicitness I give to my expression of the very same speech act. I can do the very same thing by uttering either (I) or (E), namely promising to come to your party tomorrow. While I leave it implicit in (I) – and thereby open for you to discover – that it is a promise I wish to give to you instead of, e.g., a threat or a warning, I make the role my utterance should play explicit in (E). Binkley (1979, p. 312) seems to agree, when he says:

> If one of two expressions is not an *accurate* expression of what the speaker means, while the other is, the two will have different meanings; but if one is simply more *precise* than the other, their meanings may be basically the same, especially if one is an explication of the other.

Given that this is what the Principle of Expressibility amounts to in the context of Searle's theory, we can render his principle thus:

Searle's explicit Principle of Expressibility: Whatever can be meant can be said *explicitly*.

An utterance of (I) could give expression to a promise just as well as to a threat or a warning: e.g., if a parent were to say (I) to her adolescent son in order to remind him that no excessive behavior among the party-goers will be tolerated. In the presence of appropriate circumstances, an utterance of (I) or (E) could also be meant ironically or as a joke: if, say for example, uttered in the situation sketched before not in order to reinforce a reminder that responsible behavior is requested, but to tease a child who is afraid of being embarrassed in front of his friends by his "old folks." A plethora of quite different intended meanings can plausibly be expressed in the same words. It is therefore obvious that

> [t]he meaning of a sentence does not in all cases uniquely determine what speech act is performed in a given utterance of that sentence, for a speaker may mean more than what he actually

> says, but it is always in principle possible for him to say exactly what he means. Therefore, it is in principle possible for every speech act one performs or could perform to be uniquely determined by a given sentence (or set of sentences), given the assumptions that the speaker is speaking literally and that the context is appropriate. And for these reasons a study of the meaning of sentences is not in principle distinct from a study of speech acts. Properly construed, they are the same study. Since every meaningful sentence in virtue of its meaning can be used to perform a particular speech act (or range of speech acts), and since every possible speech act can in principle be given an exact formulation in a sentence or sentences (assuming an appropriate context of utterance), the study of the meanings of sentences and the study of speech acts are not two independent studies but one study from two different points of view. (Searle, 2011, p. 18)

In consequence,

> [...] cases where the speaker does not say exactly what he means—the principal kinds of cases of which are nonliteralness, vagueness, ambiguity, and incompleteness—are not theoretically essential to linguistic communication. (Searle, 2011, p. 20)

Searle (2011, p. 57) wishes to focus on "[...] serious and literal[*] linguistic communication [...]" while excluding forms which "are not theoretically essential to linguistic communication" (Searle, 2011, p. 20), and he makes efforts to demarcate this focus:

> [*] I contrast "serious" utterances with play acting, teaching a language, reciting poems, practicing pronunciation, etc., and I contrast "literal" with metaphorical, sarcastic, etc. (Searle, 2011, p. 57, n. 1)

The justification for studying explicit and literal expressions of speech acts while neglecting metaphorical, ironical, ambiguous, etc. formulations is his Principle of Expressibility, which claims that

> [...] all what is meant (and also all what is said in a mistaken or non-literal way) can *in principle* be paraphrased by an unambiguous, explicit, literal and completely equivalent expression [...] (Kannetzky, 2001, p. 195).

This interpretation of what Searle had in mind when he introduced "exact" in his formulation of the Principle of Expressibility fits seamlessly with the methodological purpose Searle (2011) indicates for his Principle of Expressibility, and this is therefore arguably the interpretation of his principle to be preferred over Binkley's.

6.2.6.3 The Methodological Purpose of the Principle of Expressibility in Searle's Speech Act Theory

The methodological aim Searle's principle serves in the study of speech acts is to allow Searle to conduct his investigation by focusing exclusively on sentences which give explicit expression to speech acts. Searle overtly expresses this purpose again when he says:

> To study the speech acts of promising or apologizing we need only study sentences whose literal and correct utterance would constitute making a promise or issuing an apology. (Searle, 2011, p. 21)

This statement regarding speech acts of promising and apologizing can be generalized to any other kind of speech act as well. The Principle of Expressibility guarantees that for every possible speech act there is always – at least in principle – a sentence available to properly (i.e., explicitly) express the speech act in question. Since there is always a literal and explicit expression available for any possible speech act, every kind of speech act can be studied by investigating its explicit formulation.

While a description of an act can be written down, not every act can as easily be retained for further treatment as its description can. So, if the speech act itself is the intended object of study, the investigator of speech acts faces the problem of how to "conserve" the ephemeral speech act for

examination. Without a proxy to serve as a record of the object to be studied, investigating speech acts must seem rather problematic given the elusiveness of its object of study. Linguistic expressions can always easily be written down and therefore, plausibly, constitute a more easily handled object of study than acts themselves. To say it with Frank Kannetzky (2001, p. 195), "Expressions are easier to catch." Therefore, the possibility of studying speech acts by studying their (explicit) linguistic expressions constitutes a considerable methodological advantage. This is the theoretical and methodological purpose the Principle of Expressibility serves in Searle's theory. The principle bridges the gap between acts – an ephemeral and complex object to study – and their linguistic expressions, which constitute a stable and rewarding study object.[17]

6.2.6.4 Binkley's Misfire

Hence Binkley (1979) arguably misunderstands the purpose of the Principle of Expressibility in Searle's account. If it is not accuracy but explicitness Searle's principle aims at, then Binkley's critique cannot undermine the purpose of the Principle of Expressibility in Searle's theory. "[... E]xplicitness is" – as Binkley (1979, p. 311; emphasis added) himself admits – "one of the types of *precision* applied to linguistic expressions." Since precision can be determined by comparing expressions alone "[...] no reference to speaker's meaning is needed in order to discuss precision" (Binkley, 1979, p. 315) and, so, the Principle of Expressibility does not require any reference to unexpressed meanings to fulfill its purpose in the context of Searle's theory.

Binkley elaborates on the nature of explicitness as a subcategory of precision:

> We must simply remember that explicitness is just one kind of

[17] A similar position is also presented in Kannetzky (2001, p. 195) where he says that the Principle of Expressibility "[...] acts as a type of *reduction principle* that allows us to reduce the investigation of linguistic actions to the investigation of explicit [linguistic] forms [...]". Kannetzky (2002) also provides a reading of the Principle of Expressibility which is in agreement with the position defended here, but with a critical outlook on the principle. For a much more detailed account of the role the Principle of Expressibility plays in Searle's theory, see Recanati (2003).

> precision [...]. Yet whatever the measure, precision can be determined without reference to any particular speaker's meaning. Just as the precision of a clock is a property of the instrument, the precision of a sentence is a property of the language and not a measure of how well it corresponds to something else (the speaker's meaning). (Binkley, 1979, p. 315)

And Binkley immediately proceeds:

> Now we should wonder whether Searle is discussing accuracy or precision (or both) when he uses the word "exact" in the principle of expressibility. We will find that he does not distinguish the two, and that the principle is granted a specious plausibility from confusions produced thereby. (Binkley, 1979, p. 315)

Finally, Binkley (1979, p. 321) concludes that the problem

> [...] is Searle's failure to distinguish between accuracy and precision, and hence his tendency to confuse the type of situation where the speaker really does not say what he means with the type of situation described in his examples. [...] In Searle's examples, the proposed "exact" expressions have more or less the *same* meanings as the "less exact" ones. The "exact" expression is an explication of the "inexact" one [...].

This nicely agrees with examples (I) and (E), provided at the beginning of section 6.2.6.2 on page 155. As we have already seen, Binkley's accusation against Searle is mistaken since Searle means "explicit" by "exact" and can therefore safely stick to discussing a sub-kind of what Binkley calls "precision" without any need to worry about accuracy. This entirely deflects Binkley's criticism from Searle's account while the present investigation arguably remains vulnerable to Binkley's skepticism about unexpressed speaker's meaning.

Insofar as Binkley's criticism primarily builds on the dubious nature of unexpressed meanings, his criticism is rendered toothless when directed against the Principle of Expressibility *in the role it needs to fulfill in Searle's account.*

Yet the same is not necessarily the case when we consider the Principle of Expressibility in its role for the present investigation – namely in its function to help us determine the relation between what can be said and what can be thought. However, a solution to this problem for the investigation at hand – i.e., a defense of the legitimacy of the notion of unexpressed meaning – needs to wait until chapter 10. For now, the next step instead consists in deepening our understanding of what the Principle of Expressibility amounts to. For this purpose, we can first analyze the formalization Searle provided for his principle and then discuss the status of the Principle of Expressibility in more detail. This will be the aim of chapter 7.

Chapter 7

How to Interpret the Principle of Expressibility?

7.1 Formal Matters

Even the exact formulation of the Principle of Expressibility Searle (2011, p. 20) provides in formal language leaves some room for interpretation. Kannetzky (2001) provides a quite fruitful discussion of Searle's formalization of the Principle of Expressibility. He points out that, although Searle (2011) frames the Principle of Expressibility in modal terms, a standard interpretation of the possibility operator cannot be what Searle had in mind. So, according to Kannetzky, interpreting the diamond (i.e., the sign "$\Diamond$") according to possible world semantics (i.e., the standard interpretation of modal operators) would not yield the desired result. In order to see why this is supposed to be the case and whether Kannetzky is right in this regard, let us first restate Searle's (2011, p. 20) formalization of the Principle of Expressibility (in adapted notation, as it has already been provided on page 146):

$$\forall S\, \forall X \Big(S \text{ means } X \rightarrow \Diamond\, \exists E \left(E \text{ is an exact expression of } X \right) \Big)$$

Searle quantifies over speakers (S), meanings (X), and linguistic expressions (E). In plain language, the formula reads as follows: For whatever meaning any speaker wishes to convey, it is possible that there is an ade-

quate (or exact) linguistic expression for the intended meaning. Kannetzky (2001, p. 198) now raises the following objection:

> The symbolization appears strange because it is only *possible* that there is a linguistic expression for what is meant by a speaker. Then it is possible, i.e. it does not contradict the symbolization, [note 10 provided on page 166] that there is *no* exact expression of X as well. However, precisely this possibility should be ruled out according to the *colloquial version* of the principle. *"Whatever* can be meant can be said" means that there is *always* a linguistic expression. This holds for *all* possible cases. Hence, we would expect 'necessary' not mere 'possible' existence.[1] For this reason, Searle's symbolic representation of the principle is not only inadequate, it is false because it contradicts the proposition that is supposed to be represented symbolically. If whatever *can* be meant can be said, then there *must* be an expression for it. Otherwise, it cannot be meant. This is nothing else than the contrapositive of the principle of expressibility. The possibility to mean something presupposes that there is a linguistic expression of it. Generally, the possibility to intend something presupposes common forms of that content that are shared by the speaker's community.

This paragraph does not only contain a strong claim, namely that Searle's formalization is not only inadequate but actually false. It also contains a quite interesting criticism of Searle's formal symbolization of the Principle of Expressibility. So I think that this quoted passage deserves some attention and detailed analysis.

Let us start from the end and consider the last sentence of the paragraph first. The sentence "Generally, the possibility to intend something presupposes common forms of that content that are shared by the speaker's

[1] Kannetzky (2001) entails the following as note 11 at this point:

> In this case, the modality 'necessity' is, strictly speaking, superfluous because of the generalization. (Kannetzky, 2001, p. 210, n. 11)

community" (Kannetzky, 2001, p. 198) is probably an overstatement since it begs the question against the opponent of the Principle of Expressibility. Yet what Kannetzky says in the lines preceding this very last sentence of the quoted paragraph is definitely true and deserves to be mentioned explicitly.

The Principle of Expressibility – that whatever can be meant can be said – in its contrapositive form claims that, if something cannot be said, it also cannot be meant. Its contraposition is a mere corollary of the Principle of Expressibility and needs to be accepted by everyone who accepts the principle. So far, there is no problem at all with Kannetzky's statement. On the contrary, it is just right and proper to point out what accepting the Principle of Expressibility (in its usual form) amounts to by stating what it implies (in its contrapositive form). What causes discomfort is Kannetzky's addition of "generally" to introduce the last sentence of the quote. This seems to suggest (at least to me) that what follows "generally" holds true generally, i.e., without any restrictions or qualifications. If we read the sentence in this way, then it begs the question. It might be true that "the possibility to intend something presupposes common forms" of a content – I take this to mean that there need to be conventionally accepted expressions – which "are shared by the speaker's community," but only *given* that the Principle of Expressibility is true as well. If the Principle of Expressibility is not presupposed, then the claim which closes the quoted paragraph blatantly begs the question.

Since its status seems to be problematic in any case, I will not dive into the intricacies of the additionally "smuggled in" aspects of the sentence: What exactly it means for a content to have a *common form*; and why and in which regard it needs to be *shared by the speaker's community*. These points might be introduced by Kannetzky in fairness, but they significantly exceed what is presented in the canonical formulation of the Principle of Expressibility which does not mention any speaker community at all, much less that it needs to share common forms of expression of a given content. Again, while it might be fair to elaborate on the principle in this way, it needs to be shown that this is the case. These additions are certainly not trivially implied by the Principle of Expressibility on its own or its contraposition. Again, if all of that is warranted, then it only holds *if* the Principle of Expressibility is

true but not *generally*.

All of this might, however, merely be a misunderstanding of what Kannetzky meant to say by introducing his sentence with the word "generally." So, it might be a merely verbal issue of no impact whatsoever which should not be overstated.[2] There remains Kannetzky's claim that Searle's formalization contradicts the intention Searle expresses in the colloquial formulation of the Principle of Expressibility, given a standard interpretation of the modal operator.

According to Searle's formalization, whatever a speaker means, it is possible that there is an exact linguistic expression for the meaning to be expressed. Kannetzky points out that Searle's formula is compatible with the scenario that it is also possible that there is no exact linguistic expression for the meaning in question. He elaborates on this claim with an illuminating note (note 10, mentioned on page 164) to the relevant passage, where he says the following:

> $\Diamond q$ contradicts $\neg\Diamond q$, but it does not contradict $\Diamond\neg q$. Both $\Diamond q$ and $\Diamond\neg q$ can be simultaneously true, but not false at the same time. The situation that I have in mind and that is possible with respect of Searle's symbolization is, again symbolically, this: $\Diamond\,(\neg\,\exists E\,(E$ is an exact expression of $X)$ and $\Diamond\,\exists E\,(E$ is an exact expression of $X))$. Of course, the first occurrence of $\Diamond$ here means another modality than the second one. It should be read in an 'ontological' way, whereas the second occurrence of $\Diamond$ is to be read rather as 'makeability' or 'feasibility', and that is presumably Searle's interpretation of $\Diamond$. However, that is not the usual interpretation of $\Diamond$ in modal logic, and it leaves space for the problem that is discussed in this paper. (Kannetzky, 2001, p. 210, n. 10)[3]

[2] Even more so when we take into account that, to the best of my knowledge, neither Kannetzky nor I are native English speakers. A quarrel about connotative niceties of a perhaps inessential part of a formulation provided in a foreign language for the author as well as for the reader does not sound like a fruitful discussion to have, by my reckoning.

[3] The notation in this quote is adapted, since Kannetzky uses – in accordance with Searle's original formulation (see footnote 8 on page 146) – "P" instead of the diamond

I suspect that Kannetzky's formula which represents the critical situation might have suffered from inaccurate typing. At least it seems more natural for me to represent what Kannetzky announces as

$$\Big(\Diamond\,\neg\exists E\big(E \text{ is an exact expression of } X\big)\ \wedge\ \Diamond\,\exists E\big(E \text{ is an exact expression of } X\big)\Big)$$

rather than

$$\Diamond\Big(\neg\exists E\big(E \text{ is an exact expression of } X\big) \text{ and } \Diamond\,\exists E\big(E \text{ is an exact expression of } X\big)\Big)$$

The relevant difference concerns the bracketing while my usage of "$\wedge$" instead of "and" is inessential. However, I will simply presuppose system S5 for modal evaluations in this text and ignore differences which might result from evaluations in different systems of modal logic. Since both formulations are equivalent in S5, we can waive the issue as a mere matter of taste.[4]

In any case, Kannetzky's criticism is clear: Searle claims in his Principle of Expressibility that whatever a speaker wishes to express, there always is a possible expression – at least in principle – to exactly meet the speaker's intended meaning. As Kannetzky points out, Searle's formalization of the Principle of Expressibility is compatible with the possibility that there is no such expression as well. Given that this is precisely the possibility which should be excluded by the Principle of Expressibility, Searle's formalization of the principle cannot be adequate.

Searle's formulation excludes the contradictory negation of the statement that it is possible that there is an exact expression of the intended meaning – namely that it is not possible that there is an exact expression of the intended meaning – but Searle's formulation does not exclude the subcontrary opposite of the statement that it is possible that there is an exact expression of the

"$\Diamond$" to express the possibility/feasibility operator.

[4] Both formulas are equivalent in systems S5 and S4: The governing diamond in Kannetzky's formula distributes, since $(\Diamond(\varphi \wedge \psi) \leftrightarrow (\Diamond\varphi \wedge \Diamond\psi))$ is logically true, and the resulting double diamond before the second conjunct collapses. However, in system T, for example, – sometimes also called "system M" (cf. Preti, 2003, p. 14) – $\Diamond\Diamond\psi$ does not collapse to $\Diamond\psi$ (as it does in S4 and S5), and therefore Kannetzky's and my formulations are not equivalent in system T.

intended meaning – namely that it is possible that there is no exact expression of the intended meaning. So far, I entirely agree with Kannetzky, and I also agree with him that Searle's formalization should probably not be read as claiming that it is possible (in the modal logical sense of the word) that there is an exact expression of the intended meaning "[. . .] in an 'ontological' way, [. . .]" as Kannetzky (2001, p. 210, n. 10) calls it. What Searle aimed at is plausibly rather something like "[. . .] 'makeability' or 'feasibility' [. . .]" (Kannetzky, 2001, p. 210, n. 10), i.e., the ability to come up with or create an adequate expression.[5]

The difference between taking Searle's claim in the sense of (metaphysical) possibility on the one hand and in the sense of feasibility or makeability on the other hand should in general not be neglected (but see footnote 6 below). Interpreting the principle in the way Kannetzky suggests, and which I (in a certain interpretation, see footnote 6 as well) second, leaves us with a much stronger claim than the standard modal logic interpretation would give us. Many more things are metaphysically possible than actually feasible. It is, for example, plausible to assume that it is metaphysically possible to travel faster than light – given that it is not an essential feature of light that nothing is faster than it – although it clearly seems to be physically impossible, and certainly not feasible at the moment, to travel faster than light. What the Principle of Expressibility claims, according to Kannetzky, is that it is not only possible in the wide, metaphysical understanding of the notion that every possible meaning can be expressed. It is possible in a much narrower understanding of the term, namely in the sense that it can actually be achieved, *viz.*, that a suitable expression can be found for any meaning

[5] See the text surrounding footnote 1 on page 226 for additional evidence for this claim. Which kind of ability is in question at this point is not entirely clear. It might be that a *general* ability (which we already encountered on pp. 75 f) to invent or use an adequate expression is sufficient to meet Searle's demand. It might also be the case that Searle had a stronger requirement in mind, e.g., *narrow* ability (see footnote 24 on page 76) or even *wide* ability (cf. Hofmann, 2021, pp. 10 f). Given that Searle wishes to make use of the Principle of Expressibility as a methodological principle for his investigation of speech acts (see section 6.2.6.3), I take it to be a plausible assumption that Searle would prefer a quite strong notion of ability, certainly stronger than general ability, probably even wide ability.

whatsoever.[6]

However, I am not entirely convinced that the issue pointed out by Kannetzky is really problematic for Searle's formalization of the Principle of Expressibility. Does the possibility that there is no exact expression of an intended meaning, which is not excluded by Searle's formalization, really contradict what the Principle of Expressibility is supposed to amount to? Kannetzky points out that how "$\Diamond$" should be interpreted in Searle's formalization does not correspond with "[...] the usual interpretation of $^{[\cdot]}\Diamond^{[\cdot]}$ in modal logic [...]" (Kannetzky, 2001, p. 210, n. 10). As I have already said, I agree with Kannetzky in this regard. I take it as a given that Kannetzky wishes to draw our attention to the fact that the problem he highlights appears if we take "$\Diamond$" in its usual, modal logic interpretation in Searle's formalization of the Principle of Expressibility. Unfortunately, Kannetzky does not explicitly tell us what "[...] the usual interpretation of $^{[\cdot]}\Diamond^{[\cdot]}$ in modal logic [...]" (Kannetzky, 2001, p. 210, n. 10) is. I take it to be uncontroversial that the common interpretation of modal operators in modal

[6] While it remains rather unclear what kind of ability Searle had in mind, I am fairly confident that Kannetzky would agree that *general ability* should be preferred (over the stronger possible readings mentioned in the previous footnote) as a specification of what "makeability" and "feasibility" amount to. In any case, since I am inclined to read the formalization on page 163 as ranging over all *possible* meanings, I also tend to read it as ranging over any *possible* speaker. Insofar as a general ability of any possible speaker amounts to an extremely wide interpretation, I think that the ability aspect (and thereby the dependence on speaker's capacities) gets practically canceled out, so that what *language* allows is in the end crucial, not what speakers are actually able to do. Taken in this reading, which I prefer, the interpretations of the Principle of Expressibility along the lines of metaphysical possibility and makeability or feasibility (in the sense of general ability) practically coincide. Whether Kannetzky would also agree with this interpretation or whether he would prefer a stronger interpretation of the Principle of Expressibility (which depends more heavily on the ability aspect and thereby on what is actually feasible for real speakers) is not clear to me. In my preferred reading, however, the difference between the metaphysical possibility interpretation and the feasibility or makeability interpretation can *practically* be neglected, despite what I said on the facing page. In other words, my interpretation of general abilities of a speaker in this case is so lenient that the speaker could also simply drop out of the formulation. What language permits is crucial for the Principle of Expressibility in my preferred reading, not what any speaker might achieve. Only if we interpret the speaker's abilities so widely that they comprise everything language allows, we might include the speaker in the formulation of the Principle of Expressibility. Yet I think that the truth value of the Principle of Expressibility ultimately depends on what language permits, not on what speakers may or may not actually achieve. The speaker is therefore inessential for the truth or falsity of the Principle of Expressibility.

logic builds on *possible world semantics*. So, we need to settle whether the situation Kannetzky describes – which renders "[. . .] Searle's symbolic representation of the principle [. . .] not only inadequate, [but even] false because it contradicts the proposition that is supposed to be represented symbolically" (Kannetzky, 2001, p. 198) – is really problematic for Searle's formalization of the Principle of Expressibility when we interpret the diamond in the sense of possible world semantics, i.e., in its standard interpretation.

If we spell Kannetzky's critique out in terms of possible world semantics, we get the result that Searle's formulation demands that for any speaker and for any meaning, if the speaker intends the meaning, then there is a possible world where an exact expression of the intended meaning exists. What Kannetzky criticizes is that Searle's formalization allows that there also is a possible world where there is no exact expression of the intended meaning. This is the standard interpretation of what it means to say that it is possible that there is no expression for the meaning in question. While the latter formulation – that it is possible that there is no exact expression – suggests that this kind of situation is in conflict with what the Principle of Expressibility is meant to claim, this impression seems to fade if we take recourse to the former, less colloquial formulation – that there is a possible world where there is no such expression.

Now, if we ask whether it is problematic for Searle's formalization that it does not exclude the case that there is a possible world where no adequate expression for a given meaning exists, I think that we must come to the conclusion that the Principle of Expressibility should not exclude this case. The Principle of Expressibility claims that for any given meaning a speaker wishes to express, there is a possible exact expression. In other words, it is always possible to express any meaning adequately. In Searle's formalization, interpreted in terms of possible world semantics, this claim is established by saying that there is a possible world where an adequate expression exists for every meaning a speaker wishes to express. That Searle's formalization does not exclude that there are also possible worlds where no such expression exists, I think, is not a problem at all. This does not undercut the underlying intention of the Principle of Expressibility in any way.

On the contrary, I think that if this case – that there is a possible world where no adequate expression for a given meaning exists – was excluded by the Principle of Expressibility, then the principle would be false. Since we quantify over all possible worlds, we have good reason to assume that among these will also be some possible worlds where there are neither any speakers nor any expressions. If the Principle of Expressibility precluded that there is a possible world where no expression for a given meaning exists, then possible worlds where no expressions exist at all would refute the principle. The Principle of Expressibility would be disproved by possible worlds which are of no interest for the evaluation of the principle at all. For the principle to be true, there needs to be at least one possible world with the required expression. That this expression does not exist in every possible world is utterly irrelevant for the question whether the Principle of Expressibility is true. So, any attempt to modify a formalization of the Principle of Expressibility to accommodate Kannetzky's critique would, I maintain, trivially refute the principle for reasons which lie entirely outside of the "area" of relevance, i.e., beyond the possible worlds which are relevant, for the Principle of Expressibility.

I therefore think, in conclusion, that Kannetzky's claim that "[. . .] Searle's symbolic representation of the [Principle of Expressibility] is not only inadequate, [but] false because it contradicts the proposition that is supposed to be represented symbolically" (Kannetzky, 2001, p. 198),[7] is mistaken if the modal operator in his formalization is interpreted in the standard way. Had Kannetzky not stated his case in the colloquially abbreviated formulation that the principle does not exclude the possibility "[. . .] that there is *no* exact expression of X as well" (Kannetzky, 2001, p. 198), but in the more precise and explicit wording that the case not excluded by the Principle of Expressibility is that there is a possible world where there is no exact expression of the intended meaning X, then – to my mind – it would have been much easier to see that his objection is mistaken.

[7] Kannetzky (2001, p. 206) eventually settles for the position that "Searle's symbolization formulates another principle than the principle of expressibility, though a closely related one." I think, in contrast, that there is no good reason to assume that Searle provided anything but a formalization of the Principle of Expressibility when he said that he would do so.

Another way to illustrate that Kannetzky's critique is mistaken lies in translating his formalization of the situation which he thinks is problematic for Searle back into colloquial English. We already settled on page 167 that Kannetzky's formalization (i.e., the second formula on page 167) is equivalent with my formalization (i.e., the first formula on page 167).[8] So, "[t]he situation [Kannetzky has] in mind and that is possible with respect of Searle's symbolization [...]" (Kannetzky, 2001, p. 210, n. 10) simply amounts to the fact that it is *contingent* that there is an exact expression for an intended meaning since the claim that it is possible that there is an exact expression and that it is possible that there is no exact expression means that it is contingent whether there is an exact expression (cf. Girle, 2009, p. 4).[9] Searle, however, explicitly stated that "[...] where there are thoughts that cannot be expressed in a given language or in any language, it is a *contingent* fact and not a necessary truth." (Searle, 2011, p. 20; emphasis added)[10] If it is a contingent fact that there is no expression for a given meaning, then it obviously also needs to be a contingent fact whether there is an adequate expression for a given meaning. In other words, Kannetzky merely explicitly and formally stated what Searle claimed all along. Insofar as Searle stated that it is a contingent fact whether there is an expression for a given meaning before he had even provided his formalization of the Principle of Expressibility, it should not be surprising – and can hardly be problematic – that his formalization of the Principle of Expressibility is compatible with the claim that it is contingent whether there is an exact expression for an intended meaning.

Kannetzky's critique of Searle's formalization of the Principle of Expressibility is consequently unsuccessful. Notwithstanding this, Kannetzky brings up further crucial considerations regarding the Principle of Expressibility

[8] It has also been pointed out there that the two formalizations are not equivalent in every system of modal logic. All the same, since the investigation at hand is restricted to system S5 and since the formulations are equivalent in S5, we can simply call the formulas equivalent for our purposes.

[9] Girle (2009, p. 4; notation adapted) states that

> p *is contingent* translates to $(\Diamond p \wedge \Diamond \neg p)$

[10] See pp. 145 f for the quote in context.

itself, rather than its formalization. These considerations of how to interpret the Principle of Expressibility independently of its logical formalization deserve thorough examination.

7.2 The Status of the Principle of Expressibility, Part 1: Transcendentality and Normativity

The Principle of Expressibility is compatible with the fact that not every language might be equipped to provide an adequate expression for every meaning. So, the linguistic expression quantified over (E) in Searle's formalization of the Principle of Expressibility (see pp. 146 and 163) does not necessarily belong to a given, specific language, but it might belong to any language. This is because the Principle of Expressibility is not concerned with any particular language but with language *per se*.

By way of discussing this and related issues, Kannetzky draws a connection between the Principle of Expressibility and private language in the context of considering the question whether there could be meanings which cannot be expressed. At this point, Kannetzky seems to make two crucial implicit assumptions. Firstly, he seems to presuppose that, if there is inexpressible content, it must be given in a language of thought (cf. Kannetzky, 2001, p. 210, n. 13). Secondly, he seems to presuppose that any language of thought must be a private language (cf. Kannetzky, 2001, pp. 199 f). I think that both of these assumptions are plainly wrong.

7.2.1 Private Language ≠ Language of Thought

I take it to be a truism that, if meaning is given in a language of thought (in case there is a language of thought), then this meaning is linguistically framed, *viz.*, content in a language of thought is linguistic content. For all that, why should we assume that every content is linguistic content, especially if we are interested in potentially inexpressible content? I cannot think

of a good reason to exclude non-linguistic content – or at least content which is not linguistically formatted – from the discussion. Since we are trying to settle the relation between the sets M and S – i.e., the sets of what can be thought and what can be said –, a discussion of non-linguistic content must play a key role in the analysis. If it should turn out that the notion of non-linguistic content is a contradiction in terms, this would certainly change the situation dramatically. Yet I am not aware of a good argument for the claim that any content needs to be linguistic content. A mere presupposition or even stipulation to this effect would probably not blatantly beg the question against any opponent of the Principle of Expressibility, but such a presumption (that every content is inherently linguistic) would certainly leave us with a much less interesting discussion.

So, Kannetzky's first implicit presupposition is not acceptable – at least not without a solid argument for the claim that non-linguistic content can be excluded from the discussion. Kannetzky does not provide an argument to this effect, and I am unaware of a good argument for this claim. So, I deem his first assumption to be unwarranted. Where does this leave the second presumption which says that any language of thought amounts to a private language? As we have seen in section 2.2.2 (see especially footnote 19 on page 34), the position that a public language might serve as the language of thought is at least discussed. So, the assumption that every language of thought needs to be a private language is also unwarranted without further argument.

Certainly, in the context where Kannetzky's presumptions come into play, namely in discussing inexpressible content, it does not really make sense to assume that ineffable meaning could be given in a public language. Still, it needs to be noted that, from assuming a language of thought, it does not automatically follow that this language of thought will also be a private language. I therefore contend that we need to distinguish the two aforementioned implicit assumptions in Kannetzky's discussion. Identifying only one silent presupposition – to the effect that every potentially inexpressible content needs to be given in a private language (of thought) – would amount to a conflation of language of thought with private language which, I think, is a

confusion Kannetzky is guilty of committing (cf. Kannetzky, 2001, pp. 199 f and 201). Although there might be a natural tendency towards the position that these two notions coincide – *viz.*, that every private language is a language of thought and that every language of thought is private – the notions of a private language and of a language of thought still need to be distinguished and must not be confused. Even more importantly, it must not be presupposed that a private language is the only way in which (possibly) inexpressible content can be given as Kannetzky (cf. 2001, p. 207) clearly seems to do.

7.2.2 Weak and Strong Private Languages?

It is not trivial to decide how grave the consequences of these incorrect presuppositions in Kannetzky's discussion are. Since his paper is called 'The Principle of Expressibility and Private Language,' calling out this alleged confusion of private language and language of thought is probably not mere pettiness. At a certain point in the course of his analysis, Kannetzky apparently feels compelled to introduce "[. . .] the difference between a 'weak' and a 'strong' private language" (Kannetzky, 2001, p. 211, n. 16). A 'strong private language' in Kannetzky's sense of the term, I gather, is simply a "normal" private language in Wittgenstein's sense, i.e., a language that cannot be used for communication since it principally cannot be understood by anyone but the speaker. A private language in this strong sense is impossible, according to the Private Language Argument: either because such a system cannot exist or because it would – due to its privacy – not count as a language.

What then of Kannetzky's notion of a 'weak private language'? Things seem to get a little messy here: From a weak private language – or "[. . .] a 'harmless' or 'tamed' private language, [. . .]" as Kannetzky (2001, p. 211, n. 16) also calls it – it *is* possible to translate into a public language. So, whatever a weak private language is, it certainly is not a private language in Wittgenstein's sense. I am tempted to say that Kannetzky's weak private language is not a private language at all – neither a harmless or a tamed one, nor any other kind of private language. A weak private language simply is

not a private language if we restrict this notion to the discussion launched by Wittgenstein as I think we should in order to avoid terminological confusion. What Kannetzky calls a "weak private language" might still be a language of thought. However, my reading is that, due to the two aforementioned implicit presuppositions, Kannetzky maneuvered himself into an impasse that forced him to come up with the unfortunate term "weak private language." Let us take a look at the anatomy of a conceptual confusion to see where, as far as I can see, Kannetzky went astray.

In order to discuss the Principle of Expressibility in the first place, we need some kind of conceptual resource to at least address the possibility of content which cannot be expressed. Kannetzky restricts his own ability to make ineffable content a subject of discussion in two ways. First of all, Kannetzky tends to consider the Principle of Expressibility not as a hypothesis about how what can be thought (M) and what can be said (S) relate to each other (as the Principle of Expressibility is regarded here). Secondly, Kannetzky adopts the two mistaken presuppositions mentioned on page 173 and discussed in section 7.2.1.

In the present investigation, the Principle of Expressibility serves as a certain claim about how what can be said and what can be thought relate to each other. Consequentially, the Principle of Expressibility might come out true or false, depending on whether S and M stand to each other as the Principle of Expressibility claims. If option 3, as introduced on page 136 and depicted in figure 6.1 on page 140, should turn out to be the relation M and S actually stand in to each other, then the Principle of Expressibility is true. (Option 1, which would also be consistent with the Principle of Expressibility, was already ruled out in section 6.1.4.4.) If option 5 correctly describes how the two sets M and S relate to each other, then the Principle of Expressibility needs to be dismissed as false. (All other options – i.e., option 2 and option 4 – were also already dismissed in section 6.1.4.4 and section 6.1.4.1 respectively.)

So, the Principle of Expressibility is either true or false, and whether it is true or false depends on the matters of fact. This is plainly how the Principle of Expressibility is treated here. That the principle is here also

used to structure the discussion about how the sets S and M stand to each other does not change the status of the Principle of Expressibility as either correctly describing how things are or not correctly describing things how they are.[11] Yet this is not how Kannetzky seems to approach the Principle of Expressibility. He does not consider it as a declarative statement which aims at describing reality. Kannetzky (2001, p. 196) says that "[. . .] the principle of expressibility is a *normative* principle. It is constitutive for speech act theory." Furthermore, Kannetzky ascribes 'transcendental validity' to the Principle of Expressibility on the same page.

There is no need to discuss the question how exactly the truth value of a principle that has transcendental validity is determined. It seems obvious that such a principle is not evaluated according to the standard of whether it corresponds to the facts. That Kannetzky also describes the Principle of Expressibility as a normative principle suggests that the principle does not have the job of describing the world correctly. Its direction of fit – to use Searle's expression (cf. Searle, 1979d, pp. 3 ff; and Searle, 1983, pp. 7 ff; see also Anscombe, 1963, § 32, pp. 56 f) – is inverse to the direction of fit for a declarative sentence. Kannetzky also claims that the Principle of Express-

[11] That the Principle of Expressibility might be an analytic claim (see section 7.3) does not affect its status of plainly being either true or false. Other analytic sentences, such as "All bachelors are unmarried," also either correspond or do not correspond with the facts. This means that analytic sentences are true or false in the very same way as empirical sentences are. (We do not have two notions of truth: one for analytic truths and one for empirical truths. Even when we say that a claim is empirically true or analytically true, we mean that it is an empirical/analytic claim and that it is a true claim. The additions "empirically" and "analytically," although they are grammatically speaking adverbs, do not modify the truth predicate on the logical level.) Merely how their truth value is settled is different for analytic statements and for empirical statements: namely by either merely considering their meaning or by also comparing what they say with how things stand in the empirical world.

This is, by the way, not by any means an attempt to undercut Hume's Fork (cf. Morris & Brown, 2021, § 5), i.e., his distinction between relations of ideas on the one hand and matters of fact on the other hand (cf. Hume, 2007, 4.1-2, p. 18), or Leibniz's distinction between truths of reason(ing) on the one hand and truths of fact on the other hand (cf. Leibniz, 1989, p. 217, § 33) or Kant's distinction between analytic truths on the one hand and synthetic truths on the other hand (cf. Kant, 1998a, p. 130 = A6 f/B10 f). I do not nurture any iconoclastic ambitions in this regard. I merely wish to point out that if we say of a sentence (or of a proposition or whatever it is we wish to ascribe a truth value to) that it is true, it is the very same thing we say regardless of whether it is an analytic or an empirical statement we talk about.

ibility

> [...] is a precondition for a meaningful way of speaking about saying, meaning, and their relation. The principle is not a descriptive statement, hence it is not subjected to the usual truth-valuations." (Kannetzky, 2001, p. 195)

From what I said before, it should be clear that I think that the exact opposite of what Kannetzky claims is actually the case. The fact that the Principle of Expressibility serves a specific methodological purpose in Searle's system (see section 6.2.6.3) seems to have misled Kannetzky to believe that it plays a constitutive or transcendental role in the theory of speech acts or even for philosophy of language in general. This becomes evident when Kannetzky wants the Principle of Expressibility to "[...] serve as a methodological principle of the philosophy of language [...]" (Kannetzky, 2001, p. 208). Neither is in fact the case.

Speech act theory, as well as philosophy of language in general, can perfectly be practiced even if the Principle of Expressibility should turn out to be false or unacceptable in any other sense. Searle simply presupposed the Principle of Expressibility for the reasons outlined in section 6.2.6.3, i.e., to make his job of studying speech acts much easier than it would have been without a principle which bridges the gap between speech acts and sentences in order to study speech acts by studying sentences. That Searle applied the Principle of Expressibility in this special methodological way neither means that the methodological application of the principle can or should be extended beyond speech act theory to the philosophy of language in general, nor does it mean that the principle can or should exclusively be considered in its methodological application.

Kannetzky – erroneously, I maintain – identified a fundamental and extraordinary role to be filled by the Principle of Expressibility. Taking the term "transcendental" in its traditional Kantian meaning, I suppose that the Principle of Expressibility states the conditions of the possibility of (conducting) speech act theory and even of philosophy of language in general,

according to Kannetzky.[12] This, as I already said, is not the case. Neither speech act theory nor philosophy of language would break down if the Principle of Expressibility turned out to be false. That said, the study of speech acts might turn out to be considerably harder than we thought if the Principle of Expressibility is not available to serve its purpose as outlined in section 6.2.6.3.[13]

[12] It might also be the case that Kannetzky had something else in mind when he said that the Principle of Expressibility has transcendental validity. An interpretation of this claim, which arguably agrees better with Searle's approach, is that Kannetzky meant a principle which is needed for a transcendental deduction (cf. Searle, 1979a, p. viii). I am far from sure whether this might be in agreement with Kannetzky's approach even though he elaborates on what constitutes transcendental validity:

> Principles are of transcendental validity because they do not form a starting point for chains of arguments, but formulate the conditions for the meaningfulness of such arguments. In this sense, the principle of expressibility is a constraint on the philosophy of language. (Kannetzky, 2001, p. 196)

I do not feel competent to judge whether what Kannetzky means converges with what Searle (1979a, p. viii) had in mind. I think that Searle aimed at a rather traditional understanding of the term since he talks about "[. . .] a transcendental deduction of the categories [. . .]" (Searle, 1979a, p. viii). Yet Searle, to the best of my knowledge, never talks about transcendentality in connection with any discussion of the Principle of Expressibility.

However, Kannetzky (2001, p. 191) states that "[t]he aim of [his] paper is to show in what sense the principle of expressibility might be reasonable rather than to interpret what Searle 'really' meant by it." I think, in any case, that Kannetzky achieved neither of these goals, notwithstanding the fact that he provides an extremely inspiring discussion.

I should also mention that I am far from disagreeing with everything Kannetzky (2001) says. We are, for example, for all intents and purposes on the same page – I think – in holding the view that it is impossible to reduce linguistic meaning to intentionality (cf. Kannetzky, 2001, pp. 196 f) although I suspect that Kannetzky and I prefer radically different routes to argue for this claim. There is also agreement regarding our common rejection of the conduit metaphor (cf. Kannetzky, 2001, p. 197), which Kannetzky (cf. 2001, p. 210, n. 8) correctly attributes to Locke (1975, Book III). However, I find that Kannetzky exaggerates his rejection of the conduit metaphor, which brings him dangerously close to Davidson's view that non- or pre-linguistic creatures have no beliefs at all (see section 3.1) when he says: "Nothing can be said about the status of *pre-linguistic* opinions, intentions and similar 'states' that are impossible to be articulated in *any* language." (Kannetzky, 2001, p. 197) This statement can certainly be interpreted in a way which makes it almost trivially true, but I think that Kannetzky rather aimed at something which brings his position into agreement with the Quinean dictum (see p. 191). However, for an important point where I find myself in fullest agreement with Kannetzky, see section 9.1.3.

[13] The reason is that speech acts themselves would need to be studied directly if explicit expressions of speech acts were not available as a convenient proxy to study speech acts "indirectly." As mentioned in section 6.2.6.3, I take sentences to be considerably easier to study than acts. All the same I do not see why speech acts could not be studied at all if they cannot be studied via explicit expressions of said speech acts. Therefore, I think,

Be that as it may, I take "Principle of Expressibility" to be neither more nor less than the name of a certain claim which is commonly stated as follows: Whatever can be thought can also be expressed and communicated in language. In this sense, the Principle of Expressibility is not extraordinary at all. It is a sentence which attempts to describe reality as it is, just like any other hypothesis. The Principle of Expressibility is not more normative or constitutive for philosophy than the claim that a water molecule consists of one oxygen atom and two hydrogen atoms is for physics or chemistry. However, the Principle of Expressibility is certainly quite fundamental since it describes – either correctly or not – the basic facts about how our communicative capabilities – in principle – relate to our cognitive capacities. The Principle of Expressibility is indeed fundamental in this sense. It is perhaps just as fundamental as the claim that every atom – and thereby every material object – consists of protons and neutrons which form a nucleus which is orbited by electrons. This is a fundamental truth about how our reality is organized. Despite their fundamentality, the Principle of Expressibility and the two aforementioned physical truths have a quite simple and in no way extraordinary job to do: describing the world, its structure and its functioning, correctly.

This, I claim, is also the best interpretation of how Searle conceived of the Principle of Expressibility himself.[14] He simply presupposed the quite fundamental and rather plain – *viz.*, in no way extraordinary but merely quite basic – truth of the Principle of Expressibility to justify his methodological approach of studying speech acts and their performance by restricting his investigation to explicit formulations which can be used to perform the speech acts to be studied. If, of course, the truth of the Principle of Expressibility is considered as a substantial and not only methodological precondition for any further investigation, then the conceptual resources to even consider potentially ineffable content become severely crippled. It is, understandably, hard to address potentially inexpressible meaning while operating on the assump-

Kannetzky's claims regarding the Principle of Expressibility and its role for speech act theory and philosophy of language are drastically exaggerated.

[14] This is, of course, not to say that Searle thought of the Principle of Expressibility as an empirical truth. He did not: see the following section 7.3, especially on page 183.

tion that the mere possibility of such content threatens to undermine every meaningful investigation of language.[15] So, Kannetzky's point of departure for an investigation of the Principle of Expressibility seems to be ill-fated from the start. On top of this uncomfortable theoretical situation come the two misguided assumptions mentioned before:[16] The assumption that every possible content must be linguistically formatted and given in a language of thought and the assumption that every language of thought needs to be a private language.

In order to evaluate the truth value of the Principle of Expressibility at all, it is nonetheless necessary to consider possibly inexpressible content, be it only to confirm that no principally inexpressible content exists. I take this to be an investigation which does not only necessarily need to be conducted, but I also take it to be possible to carry out this inquiry in terms of a straightforward philosophical investigation. Regarding Kannetzky's investigation, the factors mentioned before converge at this point to make the task of merely considering potentially ineffable meaning hardly feasible.

The only way out of this impasse for Kannetzky, I speculate, was to come up with the unnecessary conceptual atrocity of a 'weak private language': A medium where any potentially inexpressible content can be given, in order to evaluate whether such content

(a) refutes the Principle of Expressibility because it cannot, even in principle, be communicated in a public language in an adequate way

or

(b) can, after all, at least in principle, be framed in a publicly accessible linguistic manner and therefore does not threaten the Principle of Expressibility.

[15] This is problematic even though Kannetzky seems to feel quite comfortable in this situation when he states: "There is not a pre-linguistically given and fixed intention that is only in need of (verbal) expression. Without a possible linguistic expression there is not an intention at all, but at best a quasi-physiological disposition or 'directedness'". (Kannetzky, 2001, p. 209, n. 1)

[16] Introduced on page 173 and subsequently discussed in section 7.2.1.

Yet all of these conceptual contortions are unnecessary if the wrong paths depicted before are not taken in the first place. These wrong paths are, first of all, the two mistaken assumptions about content (which does not necessarily need to be linguistically framed) and privacy (which is not a necessary feature of a language of thought) and, secondly, the superelevation of the Principle of Expressibility above a merely declarative sentence or descriptive claim.

This concludes my diagnosis of confusions and distortions in Kannetzky's – despite all of these shortcomings and wrongheaded starting points – highly fruitful paper. Nevertheless, the diagnosed problems of Kannetzky's (2001) paper converge on a position which baffles any serious evaluation of the Principle of Expressibility's truth value. However, as already mentioned, if the aforementioned mistakes are avoided, the path is clear towards a proper investigation of the question whether the Principle of Expressibility is true. At least, the path is almost clear; the accusation that talking about unexpressed meanings is theoretically impermissible is still hanging over the current investigation like a sword of Damocles. However, we will let it hang there safely until chapter 10. For now, after we have seen that the Principle of Expressibility is neither a normative nor a transcendental principle, we will consider whether it might be an analytic principle.

7.3 The Status of the Principle of Expressibility, Part 2: Analyticity

If we wish to figure out whether the Principle of Expressibility is true by determining how the sets of the "thinkable" (M) and the expressible (S) relate to each other, we arguably need to make recourse to unexpressed meanings. Since the restricted methodological use Searle makes of the Principle of Expressibility does not commit him to unexpressed meanings, he is not vulnerable to any criticism from that direction. I, in contrast, do not only discuss a strengthened version of the Principle of Expressibility due to my rejection of the possibility of a private language, which suggests that not only everything that can be meant can be said but also that everything that

can be meant can be expressed in a way which makes it understandable for someone else.[17] I also put the Principle of Expressibility to use in a very different context from Searle's. I consequentially cannot simply shrug off the clue that unexpressed meanings are mere figments of the philosopher and must therefore not be admitted in the investigation at hand.

Before we consider the issue of how questionable a recourse to unexpressed, merely intended meanings actually is, we should take a closer look at the status of the Principle of Expressibility: Is it an empirical claim which can be refuted by contradicting observations, or can the Principle of Expressibility not be disproved in this way because it represents, say, rather a normative requirement (cf. Kannetzky, 2001, p. 196) according to which nothing that cannot be properly expressed in language should be considered a proper thought?

7.3.1 Why Should We Think That the Principle of Expressibility is Analytic?

According to Searle (2011, p. 17), it is "[...] an analytic truth about language that whatever can be meant can be said." That Searle calls the Principle of Expressibility an analytic truth does not necessarily make it a normative principle, to be sure, but it would make the principle irrefutable by empirical findings nonetheless. Searle, unfortunately, does not elaborate on why he takes the Principle of Expressibility to be analytically true. The suspicion might arise that, by calling the Principle of Expressibility an analytic truth, Searle merely vents his strong commitment and confidence instead of making an evaluable claim. Jesús Navarro-Reyes (2009) fortunately steps in and provides an explanation for the puzzling claim[18] that the Principle of Expressibility is immune against empirical refutation. He writes that

> [...] no particular experiment could ever falsify [the Principle of Expressibility], because such an experiment would be hardly

[17] These matters were discussed in section 6.2.4 and in section 6.2.5.

[18] The claim is, of course, only puzzling if we take the Principle of Expressibility to (correctly) *describe*, instead of stipulate, how the sets S (what can be said) and M (what can be meant) relate to each other.

> conceivable: if I claimed that there is a particular content that could neither be expressed in fact by a particular speaker, nor in principle by anyone else, it would be a precondition for my experiment to be accepted as such, to indicate *what* is the content that the speaker is supposed not to be able to express. Would I not have to express it in order to let my hypothesis be considered by others? In that case, if that particular content is expressible *de facto*, at least for me, why shouldn't it also be expressible in principle for that speaker? [The Principle of Expressibility] seems to be proved by *reductio ad absurdum*, since there is apparently no way to formulate the alternative possibility. [Footnote omitted] (Navarro-Reyes, 2009, p. 284)

At first glance, this argument provides a sensible reason for holding the position that the Principle of Expressibility is not susceptible to (empirical) refutation. On closer inspection, however, Navarro-Reyes's defense of the analytical status of the Principle of Expressibility bears striking structural similarities with George Berkeley's (cf. 1999b, § 23, pp. 33 f) *Master Argument* and might therefore be just as questionable as Berkeley's argument. In order to elaborate on this suspicion, we need to take a quick look at Berkeley's Master Argument and compare it with Navarro-Reyes's reasoning.

7.3.2 The Master Argument

In order to prove that material objects – conceived of as non-mental entities which exist independently of the mind – cannot exist, Berkeley poses a challenge:[19] Try to imagine an object which exists unperceived, i.e., an object as existing independently of the mind. Any attempt to come up with an example which conforms to the task at hand will fail since by conceiving of such an object you will end up with a perceived object (cf. Gallois, 1974, p. 56). The very act of trying to fulfill the task at hand will be self-defeating, and

[19] A presentation of Berkeley's (cf. 1999b, § 23, pp. 33 f) Master Argument as a *challenge* can also be found in DePoe (2011, pp. 68 f). Holden (2019, p. 117; emphais added) even talks about "[...] the opening challenge that frames *each statement* of the master argument".

the impossibility of coming up with an example for an unperceived object is supposed to show that such an object – i.e., a material object – cannot exist.

The argument may need to be amended by the so-called Inconceivability Principle, i.e., a premise to the effect that what cannot be conceived of cannot exist. This principle is, however, highly implausible and will not lend additional credibility to Berkeley's argument. The Inconceivability Principle needs to be sharply distinguished from its converse, the Conceivability Principle. The latter claims that, if something can be (*ultima facie*) conceived of, then it is also possible. Together with its contraposition that something is inconceivable if it is impossible, the Conceivability Principle is much more credible than the Inconceivability Principle, which implies that everything that is possible can also be conceived.

This latter claim arguably overestimates human cognitive abilities by far. I will give just one offhand example: It is possible, I gather, that the structure of our reality corresponds to ten-dimensional spacetime, at least according to some current versions of string theory. Even higher dimensional structures, it seems, are possible.[20] My power of imagination, I must admit, quickly starts to fade when I try to imagine spatial dimensions beyond the familiar three-dimensional space.[21]

In order to turn this rather polemic skit into a solid argument, the relation between what it means to conceive of something in contrast to what it means to imagine something needs to be clarified. If, in order to conceive of something, hardly more than intentional directedness (aboutness) of a mental act is required, then only a straw man will catch fire from the critical heat of this sketched objection to the Inconceivability Principle. If, on the other hand, conceivability is more demanding and requires aptitude to be (imagistically) imagined,[22] then we might have the beginnings of an effec-

[20] According to Max Tegmark (2015, p. 149), it is even "[...] the most popular string-theory models [...]" which claim so: "the true space always has nine dimensions, but we don't notice six of them because they're microscopically curled up [...]" (Tegmark, 2015, p. 149). Nine spatial dimensions, together with one temporal dimension, gives us ten-dimensional spacetime. M-theory, to the best of my knowledge, even requires ten spatial dimensions, so we might even have eleven-dimensional spacetime.

[21] Abbott (2010) might be of some help, but he does not, of course, solve the problem.

[22] As seems to be the case for Berkeley since Holden (2019, p. 110) points out "[...] his

tive objection against Berkeley's Master Argument – at least if we take it to be based on the Inconceivability Principle. The strength and plausibility of the Inconceivability Principle, as well as of the Conceivability Principle, therefore depend (unsurprisingly) on how the notion of conceivability needs to be understood, i.e., the question what it takes to (successfully) conceive of something.

On the other hand, Thomas Holden (2019) makes a strong case for the claim that Berkeley does not accept the Inconceivability Principle in any case, and that his argumentation consequently does not rest on this principle. For Holden (2019, p. 107), it is "[...] an opinion strangely prevailing amongst Berkeley scholars" that Berkeley makes use of the Inconceivability Principle in his arguments, notwithstanding his clearly expressed opinion that there are possible things or states of affairs which elude (at least human) conceivability (cf. Holden, 2019, pp. 110, 112, 114, and 118). In consequence, Berkeley obviously cannot endorse the Inconceivability Principle. Holden (cf. 2019, pp. 108 f) suggests instead that Berkeley makes use of the much weaker and more plausible 'Contradiction Principle', according to which not inconceivability but "[...] *contradiction* entails impossibility" (Holden, 2019, p. 108).

If it is indeed the case that Berkeley rejects the Inconceivability Principle, then his modal epistemology probably comes out much more similar to David Hume's (cf. Lightner, 1997) than is usually assumed. However, since my objection to Berkeley's Master Argument – which takes the form of a *reductio ad absurdum* and will be presented on page 188 – does not take recourse to the Inconceivability Principle, these issues do not need to be solved for present purposes. It should still be noted that, even if Berkeley does not directly argue from inconceivability to impossibility, but rather from contradictoriness or inconsistency to impossibility, he still needs to demonstrate that material

[i.e., Berkeley's] apparent willingness to equate conceiving with imagining" and suggests that for Berkeley "[...] conceiving is a matter of framing ideas in the imagination" (Holden, 2019, p. 111) "[...] by actual human minds" (Holden, 2019, p. 111, n. 8). For this matter, see also Gallois (1974, pp. 59 f), who speaks of 'the imagistic criterion' in this context and Howard Robinson, who talks about 'Berkeley's imagistic theory of thought' (Berkeley, 1999a, p. 210).

objects are contradictory for his Master Argument to take off. The suspicion remains that he does so via the challenge posed in the Master Argument by showing that the notion of a material object is inconsistent because an object which exists independently of any mind cannot be conceived of. If this is Berkeley's argumentative route, then we might concede that he does not rely on the Inconceivability Principle, but this will not improve his Master Argument even one bit. At least, it will not do so as long as Berkeley cannot show that inconceivability is a better guide to inconsistency than it is to impossibility.

Be that as it may, we are at the moment neither concerned with the most historically accurate version of Berkeley's Master Argument, nor with its argumentatively strongest version. What we are looking for is a close analogy to Navarro-Reyes's proposal, and, in order to achieve this, the exegetical intricacies related to Berkeley's work are of rather marginal interest. So, we should come back to Navarro-Reyes's attempt – introduced on pp. 183 f – to explain how we can think of the Principle of Expressibility as an analytic principle.

7.3.3 The Master Argument and the Principle of Expressibility

We can stay indifferent regarding the question whether the challenge presented at the beginning of section 7.3.2, on page 184, matches Berkeley's intentions or merely represents a popularized "mythical" version of the argument. Still, we can render the argument in a way that fits Navarro-Reyes's presentation more closely: In order to refute the claim that nothing can exist unperceived, a counterexample to the claim – i.e., an unperceived object – needs to be presented.[23] But in order to present such an example, I need to perceive of the object which is supposed to fill the role of a counterexample. Any attempt to perceive of an unperceived object is therefore self-stultifying,

[23] Whether there might be other options to counter Berkeley's immaterialism is irrelevant in this context. What is relevant is that Berkeley sets up his Master Argument in a way which suggests that only presenting a counterexample could refute his position.

and no counterexample can ever be presented.

This argumentative structure matches Navarro-Reyes's suggestion: In order to refute the claim that every content can be expressed in language, a counterexample needs to be presented.[24] Yet, in order to present a counterexample – i.e., a content which cannot be (linguistically) communicated – the content in question needs to be specified in a way which admits of intersubjectively replicable communication of the content to serve as a counterexample. If a content can be framed in this way, it can obviously be linguistically expressed. So, as in the case of Berkeley's Master Argument, any attempt to refute the Principle of Expressibility must undermine itself.

However, that it is impossible to present a counterexample does not necessarily mean that there is no counterexample to the principle in question. Let us go back to Berkeley's Master Argument to evaluate whether this argumentative strategy is cogent. All Berkeley's Master Argument proves, it seems, is that I cannot think of an object without thinking of it. This is certainly true, but it hardly proves that nothing can exist mind-independently.

If Berkeley's argumentative strategy was sound, then this structurally analogous argument should be acceptable as well: If you challenge me to name a fact I do not know, I will not be able to comply with your request.[25] In order to name a fact, I must be aware of – i.e., know – the fact I come up with; I at least need to take the fact to be a fact, since otherwise I would fall short of meeting your criteria. Yet, drawing the conclusion that I am omniscient (because there is no fact I do not know) from my inability to live up to your challenge would be blatantly absurd. That I cannot name a fact I

[24] Here, again, it is irrelevant for the present context whether there might be other ways to refute the Principle of Expressibility. What counts is that Navarro-Reyes sets his case up as if only a counterexample could do the job.

[25] In order for this analogy to work we need to presuppose, of course, that naming requires providing an informative description of the fact in question which allows identification independently of the present challenge. So, coming up with a sophism like "I hereby name the first new fact I will come to know in ten minutes from now 'Bob'," in order to mention Bob as an answer to the posed challenge will not do unless Bob can be identified independently of the description which was used to fix "Bob"'s reference. Thinking back to David Kaplan's example "I am here now," mentioned in footnote 7 on page 143, we might say that I need to know *which* state of affairs is denoted (and that it obtains) by the answer I come up with to meet the challenge.

do not know does, of course, not prove that there *is* no fact I do not know.[26]

The situation is just the same with Berkeley's Master Argument: That I cannot provide an example of an unperceived object does not prove that there is no unperceived object – much less that there cannot be an unperceived object. The same should also hold true for Navarro-Reyes's reasoning: That I cannot provide an example for ineffable content – since I would need to express the content in order to provide it – does not prove that there is no content which cannot be expressed. More generally speaking, that a claim T cannot be *constructively* refuted – i.e., refuted by way of providing a counterexample to T – does not imply that T is true.[27] To claim that the Principle of Expressibility cannot be refuted because in order to refute it I would need to express inexpressible content seems to be proof by philosophical sleight of hand, rather than a legitimate reason to endorse the principle. Put differently, Navarro-Reyes's argument to show why the Principle of Expressibility is analytic uses the same trick I can use to "prove" to you that you are omniscient.

[26] A vaguely similar idea might be present in Gallois (1974, p. 65) where he says about a certain formulation of the Master Argument that "[...] we can equally derive from it a solipsistic conclusion [...]". With a slight adaptation of the challenge presented in this paragraph, we could also try to convince an interlocutor that there is nobody whom she does not know – if this is not already implied by her omniscience. If our inability to come up with an example of an unperceived object warrants the conclusion that there is no object which exists mind-independently, then it should also be possible to extend this conclusion to the domain of other people. Yet, Berkeley was of course far from subscribing to solipsism. Spirits (or souls, or minds), according to his system, constantly perceive themselves and are therefore not in need of being perceived by any other mind to fulfill Berkeley's criterion for existence – *esse est percipi* – that to exist is to be perceived. I thus think that Berkeley has adequate resources to block a derivation of solipsism from his Master Argument without any need to take recourse to ad hoc measures. It is hard to decide whether Gallois's suggestion regarding solipsism in fact follows a similar idea as the *reductio* presented here, due to the extreme terseness of the passage in question (cf. Gallois, 1974, p. 65). If he had something similar in mind, his example, I think, lacks cogency – for the special status of perceiving entities, i.e., minds, in Berkeley's philosophy mentioned before. But as of yet I have not in literature come across a better match to the *reductio* of the Master Argument that I presented here.

[27] I owe this formulation to Frank Hofmann.

7.3.4 Summary

For the sake of clarity, I think it is in order to quickly recapitulate the argumentative structure of this section: I claim that Navarro-Reyes's proposal to explain why the Principle of Expressibility is analytic bears striking structural similarities with Berkeley's Master Argument. I further claim that, if Berkeley's Master Argument was sound, then we could demonstrate to any given interlocutor that she is omniscient. Since an argument which is structurally equivalent with Berkeley's Master Argument leads us to an absurd conclusion, the Master Argument must also be structurally flawed.

Let me again emphasize that any historical inaccuracy in my depiction of the Master Argument is beside the point because my argumentative aim is not to actually refute Berkeley. The relevant point is that the version of the Master Argument discussed here – even if it might just be a straw man version of Berkeley's actual argument – bears sufficient structural similarity with Navarro-Reyes's suggestion to rule out his proposal in the same way in which the Master Argument can be refuted, namely via *reductio ad absurdum*. That the Principle of Expressibility is an analytical truth therefore cannot be shown in the way Navarro-Reyes (cf. 2009, p. 284) proposes.[28]

However, a demonstration that a certain proposal about how to think of the Principle of Expressibility as an analytic principle fails does not, of course, show that the Principle of Expressibility is not analytic. We will come back to the question whether the Principle of Expressibility is an analytic principle in section 8.4. First, I wish to come back to the issue regarding the question whether (potentially) ineffable content is a theoretically legitimate notion. I take this question to be of central importance for the present investigation, and I take it to be a question which is not easily answered. The following chapter 8 will therefore merely serve as a preparation for my final take on unexpressed meaning, which will be presented in chapter 10.

[28] Note that Navarro-Reyes (2009) does not actually commit himself to the truth (or analyticity) of the Principle of Expressibility on the basis of his "demonstration" of the principle's analyticity. If certain formulations in this section have suggested otherwise, this is due to my attempt to keep the discussion as simple and straightforward as possible by avoiding more cumbersome formulations which would have probably done more justice to Navarro-Reyes's position by mirroring his rather critical perspective on the principle.

Chapter 8

Is It Legitimate to Speak Of Unexpressed Meanings or Even Ineffable Content?

Preparatory Remarks

Let us come back to the question whether unexpressed meanings are eligible theoretical entities. I will not try to decide whether unexpressed meanings should be thought of as mental or rather as abstract entities, e.g., propositions. The following considerations will be applicable, I think, to both options. As already mentioned, a rejection of unexpressed meanings is often based on Quine's famous dictum: "No entity without identity[!]" (Quine, 1969, p. 23) As this slogan is commonly understood, it prescribes that nothing – i.e., no kind of thing – should be admitted in one's ontology if one is not able (or willing) to provide precise identity and individuation criteria for instances of the kind of object in question.[1]

[1] This position seems to be foreshadowed already when Gottlob Frege writes to Edmund Husserl on December 9, 1906: "It seems to me necessary to have an objective criterion for recognizing a thought as being the same, because in the absence of such a criterion logical analysis is not possible." (Mohanty et al., 1974, p. 91)

8.1 The Problem With Verificationism

In its crudest form, this prescription is probably a consequence of exaggerated verificationist scruples. The verificationist theory of meaning can be cited for explanation and illustration. According to the verificationist theory of meaning, "[...] the meaning of a statement lies in the method of its verification." (Carnap, 1966, p. 76)[2] It follows from this claim that, if there is no method of verification, then there is no meaning. To put it in the words of Rudolf Carnap, one of the verificationist movement's most acute thinkers: A word or sentence "[...] remains meaningless as long as no method of verification can be described." (Carnap, 1966, p. 66)[3] This, we may say, is an extreme exaggeration of empiricism which is founded on a confusion of definition with criteria. Let us take one thing at a time.

Why is the verificationist theory of meaning mistaken? The reason is not that truth and meaning are not related. They are,[4] but not in the epistemically charged way envisaged by verificationism. The reason is also not that a verificationist theory of meaning cannot handle anything but declarative sentences. It is obvious that sentences like "Will you please come to the party tomorrow?" and "Don't spit on the floor!" clearly have meaning,

[2] In the German original, Carnap (1931, p. 236) says "[...] daß der Sinn eines Satzes in der Methode seiner Verifikation liegt."

[3] In the German original, Carnap says that any word or sentence "[...] bleibt auch weiter bedeutungslos, solange man keinen Weg zur Verifikation angeben kann." (Carnap, 1931, p. 225)

[4] I base this claim on a strong sympathy towards truth-conditional semantics (cf. Wittgenstein, 2001; 2003a, § 4.024), which will be merely stated but not defended here. Truth-conditional semantics, however, can be charged with the very same criticism which will be quickly mentioned in a moment as directed against verificationist theories of meaning: namely that they ignore most of language which does not fit in the category of descriptive sentences. In contrast to verificationist theories, truth-conditional accounts can be easily generalized to incorporate *fulfillment conditions*, which allows a treatment of non-descriptive statements as well. Whether the same can be done for verificationism – *viz.*, whether something like "methods of fulfillment" can be adopted in analogy with methods of verification – seems less clear to me. But this is, as already mentioned, not the place to argue for the advantages of truth-conditional semantics over verificationist semantics. Also, my main argument against verificationism (to be presented in section 8.3.1) is, to my mind, independent of any potential amendment of verificationism with "methods of fulfillment" – to the extent that this notion makes sense at all – and the question whether verificationism can be saved via this route therefore does not need to be answered in this context.

although they cannot be "verified." This is a fair objection to make against a verificationist theory of meaning, but it is not the objection I have in mind.

The problem with the verificationist theory of meaning is not that it is not applicable to most of language (given that most of language is not declarative). The problem is that a verificationist theory of meaning is not even suitable to handle its proper domain of application, namely declarative sentences which are meant to (correctly) describe reality. The reason why a verificationist theory of meaning is not even adequate for its own proper domain is, I hold, a deeply engrained tendency to mix up (and in consequence confuse) philosophical domains which need to be kept separate for lack of good reason to admit reciprocal effects of one domain on the other:[5] metaphysics and semantics on the one hand, and epistemology on the other hand. The former deals with metaphysical definitions (what the world is like) while the latter is primarily concerned with criteria (how we come to be justified in applying said definitions). To frame the contrast in different terms, we may also take recourse to John Locke's distinction between 'real and nominal essence' (cf. Locke, 1975, III.iii.15 ff, pp. 417 ff), where real essence is merely concerned with what (the nature of) a thing is while nominal essence provides criteria to recognize something as being of this or that kind (cf. Jones, 2018).

8.2 An Illustration by Way of Theories of Truth

It might sound odd, at first glance, to lump metaphysics and definitions together, and contrast both with epistemology and criteria. So, let me explain what I mean by way of an example: theories of truth. An ideal theory of truth would not only provide us with a metaphysical definition of what (the nature of) truth is in terms of necessary and sufficient conditions; it would also offer criteria for us to separate truths from falsehoods. Unfortunately, we do not have such an ideal theory of truth. What we have are several

[5] This point echoes and builds on Saul Kripke's admonition not to conflate different domains of philosophical investigation; see page 211 below for some elaboration. See also Kripke (1981, p. 49).

theories which accomplish different tasks we would like to see unified in an ideal theory of truth.

Let me elaborate: The correspondence theory of truth – claiming that truth consists in agreement with the facts, or in representing things as they are – gives us, I claim, a perfectly cogent *definition* of what truth is. Unfortunately, this definition is of no help whatsoever when we try to find out what is true and what is not, i.e., when we wish to distinguish truths from falsehoods. This means that the correspondence theory of truth does not provide any *criterion* we could make use of to identify truths.

Other truth theories were suggested to fulfill this task, e.g., coherence theories of truth, pragmatist theories of truth, or the consensus theory of truth.[6] They all provide – in contrast to the correspondence theory of truth – criteria which can help us to sort out truths and falsehoods. These and other truth theories are, for good reason, often collectively labeled 'epistemic theories of truth,' (cf. David, 2016, § 8.1) and they can be contrasted with a realist theory of truth like the correspondence theory. Epistemic theories of truth have plausible and valuable contributions to make if what they suggest is taken as a *criterion* to find truth. If, however, these theories are understood as proposing alternative *definitions* of truth, they will need to be dismissed as incorrect. So, as long as the different contributions to the debate about truth take their appropriate places (i.e., the correspondence theory as dealing with a metaphysical definition of truth and the epistemic theories of truth as being concerned with criteria to find truth) the several truth theories do not even necessarily stand in conflict with each other but might, on the contrary, collaborate to provide a fuller picture of how to acquire what the correspondence theory specifies.

I have no space to argue for my assessment of the situation – that correspondence provides the (metaphysical or "real") definition of what truth is while other theories provide criteria for how to find truth – here, but, despite the crudeness of the picture I just sketched of the situation regarding

[6] I speak of "*the* correspondence theory of truth" and "*the* consensus theory of truth" for mere convenience. They come in various versions, just like coherence theories of truth and pragmatist theories of truth. So, I suppose, it would be just as legitimate to talk about the former two in the plural as it would be to talk about the latter two in the singular.

(apparently) conflicting theories of truth, I am convinced that an approach along these lines is on the right track. These few lines will, of course, not be sufficient to solve the age-old philosophical debate about truth. Even so, they should sufficiently illustrate why I contrast metaphysics, including metaphysical definitions, on the one hand with epistemology and criteria on the other hand and why I take it to be of utmost importance not to confuse or conflate these domains. With this illustration at hand, let us come back to the questions of how and why exactly a conflation of metaphysics/definitions with epistemology/criteria is problematic in the case of verificationism.

According to a solid realist approach to metaphysics and epistemology – i.e., what I take to be simply *not* confusing the two – it is obvious that reality is independent from how and what we think about it. This, of course, also includes that what is true and what is not true are generally independent of what we think is true or not. Also what is true and what is not true are independent of our epistemic access to – i.e., our prospects of finding out – what is true and what is not true. Given that neither are we made to understand reality, nor is reality made to be understood by us, there might plausibly be truths we cannot discover. I am therefore prepared to accept "[...] the principle of transcendence (which says that a proposition may be true even though it cannot be known to be true)" (Young, 2018, § 1).[7] At least, it seems that this is an option which cannot simply be stipulated away. Yet, this is exactly what verificationism does by banishing everything which cannot be verified from the domain of the meaningful.

8.3 Refuting Verificationism

Verificationism is false, I claim. But it is not trivial to find an obvious counterexample against the verificationist theory of meaning, for even the most rudimentarily refined versions of this theory will insist that only what cannot be verified even *in principle* is meaningless. The plausibility of any

[7] I should emphasize again at this point that while I accept a principle of transcendence with regards to truth, i.e., correctness$_t$, I do not accept a corresponding principle of transcendence with regards to correctness$_i$, i.e., inferential correctness.

given variant of a verificationist theory of meaning will therefore depend on how "in principle" is spelled out in this context. The more lenient and inclusive "in principle" is interpreted, the harder it will be to find a declarative statement which is plausibly meaningful but nevertheless not verifiable, even in principle – in other words, a counterexample to the verificationist theory of meaning.

The statement that Caesar had scrambled eggs for breakfast on his twelfth birthday is certainly meaningful, but whether it can be verified depends on the method of verification permitted by the hedge phrase "in principle." If verifiability in principle admits of traveling back in time to check what young Caesar had for breakfast – notwithstanding the fact that we cannot actually do that and that time travel might even be impossible – then the sentence "Caesar had scrambled eggs for breakfast on his twelfth birthday" will come out as meaningful, even according to the verificationist theory of meaning.

John Austin, at the beginning of his 'Performative Utterances,' rightly points out that verificationism, or 'the verification movement' (Austin, 1970, p. 234), as he calls it, "[...] did a great deal of good; a great many things which probably are nonsense were found to be such." (Austin, 1970, pp. 233 f) I definitely agree. Thanks to the rigorous criticism of metaphysical antics, accomplished by verificationist thinkers, issues such as "is the number 7 holy?" or "which numbers are darker, the even or the odd ones?" (Carnap, 1966, p. 72) were relentlessly unmasked as nonsensical pseudo-questions (cf. Carnap, 1966, p. 72). Carnap's (cf. 1966, pp. 69-72)[8] "deconstruction"[9] of Heidegger's claim that *"The Nothing itself nothings."*[10] can be mentioned as a prominent example of verificationist endeavors to expel metaphysics gone astray. While that was then, things certainly seem to stand a little differently

[8] The corresponding passages in the German original can be found in Carnap (1931, pp. 229-232).

[9] Not in, e.g., Jacques Derrida's sense of the term or any other postmodernist understanding, of course.

[10] This translation of Heidegger's (1955, p. 34) "Das Nichts selbst nichtet." appears in Carnap (1966, p. 69). In an English edition of Heidegger's texts, however, the relevant passage is rendered as saying "Nothing 'nihilates' (*nichtet*) of itself." (Heidegger, 1949, p. 369) I will not attempt to judge which translation may be more faithful to the original since I agree with Carnap that the sentence in question does not mean very much either way.

nowadays.

Austin goes on to say that "[. . .] perhaps some things have been dismissed as nonsense which really are not; but still this movement, the verification movement, was, in its way, excellent." (Austin, 1970, p. 234) Again, I tend to agree, but it should be noted that Austin states the matter in a rather polite way. Instead of speaking of *some* things which have *perhaps* been erroneously dismissed as being nonsensical, we might also say that verificationism threw out the baby with the bath water.

8.3.1 Verificationism and Goldbach's Conjecture

As a telling problem case for the verificationist theory of meaning, we can take a look at *Goldbach's conjecture*. The conjecture was proposed by Christian Goldbach to Leonhard Euler in 1742 and, notwithstanding several attempts, remains unproved up to the present day (cf. Vaughan, 2016, p. 479). The (binary) Goldbach conjecture, rendered in modern terms, claims that "[e]very even integer greater than 2 can be written as the sum of two primes." (Vaughan, 2016, p. 480) Since this conjecture has neither been proved nor refuted yet, we neither know its truth value nor how to prove it. Otherwise it would have been proven already.

The Goldbach conjecture is certainly uncomfortable terrain for the verificationist, but to draw the curtain over the verificationist theory of meaning already would be premature.[11] Let us take a closer look and see how the

[11] It needs to be noted that a verificationist might wish to reject the challenge coming from Goldbach's conjecture right away because the Goldbach conjecture is a mathematical claim while the principle of verification – as stated on page 192 – only applies to empirical statements. However, we should not let the verificationist off the hook so easily. Arguing that Goldbach's conjecture is analytical, and therefore not subject to the verification principle, does not seem all that convincing since Goldbach's conjecture does not appear to be true by virtue of its meaning alone. The Goldbach conjecture should be read as an attempt to state a mathematical fact, not a semantic fact or a stipulation. Although the Goldbach conjecture is certainly not an empirical claim, it also does not seem to be a tautology or a contradiction. Pairing Goldbach's conjecture with logical truths/falsehoods to put it out of reach for the verification principle therefore also does not seem to be fully convincing. A more promising approach for the verificationist to safeguard mathematical claims appears to be the pronouncement that there is no noteworthy problem regarding the verifiability of mathematical statements since they can be "calculated" to determine their truth value, i.e., it simply takes calculation to determine whether "$2 + 2 = 4$" is true.

verificationist might accommodate Goldbach's conjecture. The challenge for the verificationist is clear: We obviously understand the Goldbach conjecture since we understand what it means that every even integer greater than two is the sum of two prime numbers. At any rate, understanding a sentence, according to the verificationist theory of meaning, amounts to having a method of verification at hand for the sentence in question. If we had a proof of Goldbach's conjecture, we would know how to verify it. Since we do not know how to prove the conjecture, the verificationist needs to explain why we understand the conjecture perfectly well nevertheless.

Which means does the verificationist have at her disposal to answer the challenge? One way out of the conundrum for the verificationist is to insist that we can describe a method of verification after all: "Go through all the even integers greater than two and check for every instance whether there are two primes which sum up to the number in question" should do. The shortcoming of this suggestion is, of course, that it cannot be carried out. Since there are infinitely many even integers greater than two and since it is impossible to actually go through an infinite series, nobody can check whether Goldbach's conjecture is correct by following this "method of verification."

The verificationist could try to stand her ground and point out that the verificationist criterion of meaning does not require an actually realizable method of verification in order to admit a sentence as being meaningful. This, I suppose, is technically correct. Yet one can, I think, still legitimately wonder: What worth can be attributed to a "method of verification" about which it is definitely known that following it will never yield a decision?[12] In order to prove Goldbach's conjecture this way, one would need to finish off checking an infinite series of numbers. As it is impossible to ever finish an infinite task, actually proving Goldbach's conjecture in this way is also

Then again exactly this claim is questionable regarding Goldbach's conjecture: It might be the case that there is a way to calculate whether Goldbach's conjecture is true or false, but we do not know what this way looks like.

[12] This formulation is not perfectly accurate, for it might be possible to achieve a result after all. Goldbach's conjecture can never be proven in this way, but following the method sketched above might be a viable procedure for refuting the conjecture nevertheless. We will come back to this fact and see whether it can be of any help for the verificationist on pp. 199 f.

impossible. Does an impossible to carry out method of verification really deserve to be called a "method of verification"? I am unsure, to say the least, but I will leave it to the verificationist to answer this question.

All the same, it can perhaps be granted that "Caesar had scrambled eggs for breakfast on his twelfth birthday" is meaningful because we could travel back in time to take a look even if traveling back in time might be impossible. It seems to be at least unclear (yet) whether time travel is really impossible.[13] Let us give the verificationist the benefit of doubt in this case. Alternatively, we might also come up with "counterfactual methods of verification" for this instance. Had anybody noted down what Caesar had had for breakfast on his twelfth birthday and had this document been conserved, we could have consulted it to decide whether the sentence "Caesar had scrambled eggs for breakfast on his twelfth birthday" is true. We do not have such a document, and, to the best of my knowledge, no such document has ever existed. Nonetheless, it is at least a perfectly conceivable situation which would allow us to decide the matter in question. For a method of verification, this might be good enough. In the case of Goldbach's conjecture, however, we definitely know that it is impossible to verify the claim in question by following the method described. Interpreting verifiability *in principle* so leniently that even a demonstrably impossible to carry out proof is admitted is probably too much of a stretch for the small hedge phrase "in principle."

Be that as it may, the verificationist does not need to try to "brute force" her way out of the challenge posed by the fact that we can perfectly understand Goldbach's conjecture without having an obviously acceptable method of verification at hand. A more clever and elegant path seems to be available as well. A *direct* method of verification for Goldbach's conjecture might not be necessary after all. Could an indirect method of verification, i.e., a method of verification for the negation of Goldbach's conjecture, maybe save the day for the verificationist?

Refuting Goldbach's conjecture seems to be theoretically and method-

[13] The kind of impossibility in question here is physical possibility or technological feasibility. It is a plausible assumption that time travel is, at least, epistemically possible, and time travel might even be metaphysically possible.

ologically much easier than proving it. Importantly, we have a clear idea of what a refutation of Goldbach's conjecture would look like. We simply need to find a counterexample to the conjecture, i.e., an even integer greater than two which cannot be written as the sum of two primes. That no such number has been found yet, and that it seems highly improbable that one will ever be found, is irrelevant. The important point is that we know of a method to verify the negation of Goldbach's conjecture. And this method of verification is at least not *guaranteed* to fail.

Is that not wonderful news for the troubled verificationist? The solution seems to be simple: Just do not insist on verification alone, but also admit of falsification (cf. Beaney, 2010, p. 811). The verification principle can easily be amended. Instead of the original formulation, given on page 192, we can alternatively render the core claim of the verificationist theory of meaning thus:

> The meaning of a statement lies in the method of its verification *or of its falsification* (i.e., the verification of its negation).

For the dogmatic verificationist who thinks that this reformulation reeks of Karl Popper, we can even avoid any reference to falsification and stick to verification exclusively:

> A statement remains meaningless as long as no method of verification *for it or its negation* can be described.

This should satisfy even the most die-hard verificationist. Unfortunately, this does not solve the problem. It might be the case that the solution sketched before immediately came to the mind of our clever verificationist. Yet I have to admit that it took me a while to come up with this solution when I first thought about the problem. Nevertheless, I understood Goldbach's conjecture right away. That is to say, I understood what it means that every even integer greater than two is the sum of two primes before I had thought of the indirect verification method for the Goldbach conjecture. Furthermore, I cannot see how I could have ever come up with this solution had I not grasped the meaning of Goldbach's conjecture beforehand. Understanding

what the sentence "every even integer greater than two is the sum of two primes" means is undoubtedly a precondition of concluding that this claim can be refuted by finding an even integer greater than two which is not the sum of two prime numbers.

This clearly shows that the verificationist theory of meaning gets the order of explanation wrong. Knowing a method of verification (or falsification) cannot be required to understand what a sentence means since I need to understand what a sentence means in order to come up with a method of verification (or falsification) for the sentence in question. It is certainly plausible that, in order to understand a sentence, I need to know what would be the case if the sentence was true – *viz.*, understanding a sentence amounts to knowing its truth conditions.[14] So, I am quite confident that truth-conditional semantics is a promising account. But taking the additional step towards verificationism by claiming that in order to understand a sentence I need to know how to find out whether the sentence is true (or false) is definitely wrongheaded because it reverses the logical order of things: First I need to understand what a sentence means, and only then can I wonder how to determine its truth value.[15]

So we can say that verificationism puts the cart before the horse, not to mention the fact that the verification criterion of meaning cannot live up to its own standard. What might the method of verification for the claim *that the meaning of a statement lies in the method of its verification* be? If we cannot provide a method of verification for this claim, then the verificationist criterion of meaning is – according to the verificationist theory of meaning – meaningless. The verificationist criterion of meaning has a self-application problem. So, on top of its inadequacy, the verificationist theory of meaning is also self-defeating.[16]

[14] As far as declarative sentences are concerned. Regarding non-declarative sentences, see the remarks in footnote 4 on page 192.

[15] This is the *hysteron proteron* objection, commonly attributed to Isaiah Berlin (1939, p. 228), but already formulated and discussed previously (cf. Uebel, 2019, p. 11, n. 40).

[16] Not that the verificationist does not have any means to answer the criticisms presented here; see Uebel (2019) for a very illuminating discussion which emphasizes that we need to distinguish 'V-CRIT' (i.e., verificationism as a criterion of meaningfulness) from 'V-TOM' (i.e., verificationism as a theory of meaning) in order to plausibly assess the verificationist

8.3.2 A Look Back at the Private Reasoning Argument

Now, after this rigorous criticism and rejection of verificationism on the basis of the accusation that verificationism confuses metaphysics and epistemology, the following concern might readily appear: Is not the same mistake also committed in this investigation? Does the author (i.e., I) not also confuse metaphysics and epistemology by restricting the domain of reasoning on the basis of epistemic considerations, namely the ability to double-check on the correctness of reasoning? I can see two, apparently conflicting ways to answer to this suspicion, and I will discuss them in turn.

8.3.2.1 First Defense

The first reply consists in pointing out that the domain of reasoning was not restricted on the basis of epistemic considerations at all. I take it to be a *metaphysical* truth about (the nature of) reasoning that it needs to be possible – at least in principle – to check whether any particular instance of reasoning is correct or not. Since the restriction follows from the nature of reasoning, we are not confronted with an epistemically motivated restriction of metaphysics but with a metaphysically motivated restriction of metaphysics, so to speak. It is considerations about the very nature of reasoning (and thereby metaphysical considerations) which demand that nothing which cannot at least in principle be made available for checking its correctness can count as reasoning. Epistemic considerations regarding the question of what we can or cannot know, or regarding what and how we can or cannot find out, are not involved in the claim that there cannot be (necessarily) private reasoning. It is the nature of reasoning itself which requires reasoning to be publicly accessible (at least in principle). It is not our epistemic constitution which makes it so.

criterion of meaning. Even so, taken together, the *hysteron proteron* objection and the self-refutation objection might be devastating enough to actually refute verificationism since what is left over after directing the *hysteron proteron* objection against V-TOM can plausibly be purged by the self-refutation objection to rule out V-CRIT. However, the subtleties regarding which variations of verificationism are prone or immune to which arguments exceed the scope of the present investigation.

The reason why the borders between metaphysics and epistemology may seem to be blurred in the Private Reasoning Argument is simply due to the topic we are concerned with. Since reasoning is one of our most salient means to acquire, and, even more importantly, to justify knowledge, there is an obvious and strong connection between reasoning and epistemology in general. Still, we should not be fooled into believing that every consideration regarding reasoning must be an epistemic endeavor by the mere proximity of the topics of reasoning and epistemology. An investigation into the nature of knowledge or justification – *viz.*, the search for an adequate definition of what knowledge or justification are – should be counted as a metaphysical investigation as well, notwithstanding the fact that it is a metaphysical investigation into the most central notions of epistemology.

One way to see that we should count such investigations as metaphysical investigations can be achieved by making use of the distinction between (metaphysical) definition and criteria again. Let us say that we have a correct (metaphysical) definition of knowledge (or justification). With such a definition at hand, it does not make sense to further ask: "But what *is* knowledge (or justification) really?" The correct definition of knowledge (or justification) is the answer to just this question and puts an end to further metaphysical inquiry – at least if the answer is accepted as the correct metaphysical definition of knowledge (or justification). The questions "How do we recognize knowledge (or justification)?", "How do we know that we know (or are justified in) something?", and "How do we know that this is the correct definition of knowledge (or justification)?" definitely still make sense, even after a metaphysical investigation is successfully brought to an end.

The reason is simple: These latter questions are epistemic questions, which remain, naturally enough, unanswered by a metaphysical investigation. The correct metaphysical definition, plausibly, only provides a "mere" definition of the phenomenon investigated without providing any criteria to recognize the phenomenon in question (as in the case of the definition of truth as correspondence with the facts). It might of course be that we even find an "ideal" theory which does not only provide the correct metaphysical definition but, by the same token, also defines the phenomenon we are

concerned with in a way which allows perfect recognition and distinction of the phenomenon in question from other, perhaps similar phenomena. In any case, I would not hold my breath for such an operational definition – as it is sometimes called when the necessary and sufficient conditions also provide criteria for recognizing what falls under the definition and what not – to be found for most of the interesting philosophical terms.[17]

I would hope that this first reply is acceptable for everyone who might suspect that I apply a double standard when I condemn verificationism for conflating metaphysics with epistemology but, at the same time, claim that reasoning is necessarily evaluable for correctness. This might, admittedly, look like an epistemic restriction on the notion of reasoning, at least at first glance. Nevertheless, the first response should, I hope, have shown that making the claim that private reasoning is impossible does not amount to a conflation of metaphysical (what reasoning is) with epistemic (how we assess reasoning) considerations. This first response has the dear advantage of keeping metaphysics and epistemology strictly separated as I like them to be. All the same, I can see several objections looming in the background, coming from those who are not willing to accept the borders between metaphysics and epistemology as I have drawn them. I take this to be a generally fair complaint since I admit that there are other reasonable ways of demarcating metaphysics from epistemology and *vice versa*. I am therefore willing to provide a second, last-resort reply to justify my Private Reasoning Argument, as well as my critique against (traditional) verificationism.

8.3.2.2 Second Defense

For the second reply, I have already conceded that the Private Reasoning Argument commits me to a mild and restricted version of verificationism. I have also already defended this mild and restricted form of verificationism in section 3.1, but this is a perfect occasion to shed some additional light on the reasons why I think that my mild and restricted verificationism is viable whereas traditional and other more radical forms of verificationism

[17] Immanuel Kant (cf. 1998a, p. 197; 1998b, pp. 136 f = A58 f/B82 f) for example suggests that such an "ideal" theory of truth is not even possible.

are untenable.

Let us, for the sake of argument, assume that the ultimate clarifications of what exactly metaphysics and epistemology are come out in a way which makes it clear that I do indeed draw a metaphysical conclusion (that there cannot be private reasoning) from epistemological considerations (that reasoning needs to be assessable for correctness) in *PRA*. I am not sure what these definitions of epistemology and metaphysics would need to look like, but so be it. I will grant, for the moment, that I have transgressed the boundaries between metaphysics and epistemology in my argumentation. How could I get out of this tight spot? I think that the following could be a way out:

I take it to be a plausible assumption that a human capacity (reasoning) is dependent on another human capacity, namely our ability to check on the correctness of reasoning, and by the same token hold that it is justified to assume that reasoning is restricted by (allegedly) epistemic considerations. I also argue, in contrast, that it is not justified to think of reality in general as being dependent on epistemology or our epistemic endowment. The latter is an unwarranted invasion of metaphysics by epistemology. The former, since reasoning is (arguably) not mind-independent, is not guilty of a confusion of metaphysics and epistemology.

This defense can also be applied to the advantage of *PLA* (in addition to the defense of *PRA* I just sketched), in case similar doubts as those I sketched at the beginning of section 8.3.2 on pp. 202 f might appear regarding the Private Language Argument as well. Since language, being a cultural good, is arguably not entirely mind-independent either (just like reasoning), drawing metaphysical conclusions about language (namely its necessarily being public) from seemingly epistemic considerations (about what it takes to follow a rule or about how rule-following needs to be achieved) can plausibly be justified also in this special case.

Nevertheless, rather than endorsing this second reply, I would prefer to take a position along the lines of the first response (from section 8.3.2.1) here as well: *PLA* illuminates the nature of language by building on insights regarding what (the application of) a rule *is* or what it amounts to if a rule is being followed. Insofar as these can be considered as purely metaphysical

considerations, there is no blending of metaphysics and epistemology going on in *PLA*. However, if this way of distinguishing the two relevant philosophical domains – namely metaphysics and epistemology – is not accepted, there is still the second reply available to explain why the interference is nevertheless permissible in this special case.

8.3.2.3 Generalizing the Second Defense: Solid, Liberally Conservative, Metaphysical Realism

In more general terms, this means that epistemic considerations can legitimately interfere with metaphysical investigations of *mind-dependent* phenomena. Nonetheless, drawing metaphysical conclusions from epistemic considerations is unwarranted if the investigation in question is concerned with mind-*in*dependent aspects of reality. This concession leaves a solid metaphysical outlook intact while it still explains why – under specific conditions – the demarcation between metaphysical and epistemic considerations is not as watertight as it generally is, according to my account.

This explains why a highly specific form of verificationism – namely the mild and restricted version I sketched towards the end of section 3.1 – does not confuse or illegitimately conjoin different philosophical domains. For, if an interaction between metaphysical and epistemic considerations is admitted, this interaction is backed up by an independent argument for the legitimacy of the case in question. Yet a general and unrestricted impact of epistemology on metaphysical considerations is not acknowledged. Rather, for every suspected interference of epistemology with metaphysics, there needs to be a cogent argument for the permissibility of a given interaction of metaphysical and epistemic considerations on a case by case basis. In order to have a label at hand, let us call this concession to possibly warranted interactions between metaphysics and epistemology "liberally conservative (metaphysical) realism": or more shortly, "liberal (metaphysical) realism" since the conservative aspect is, in a way, expressed in "realism" already.

This liberal realism is consistent with mild and restricted verificationism, and it is also consistent with solid metaphysical realism, according to which

reality is independent of our take on it. This includes a commitment to the *principle of transcendence* which claims that there might be (true or false) propositions which lie outside of our epistemic reach (cf. Young, 2018, § 1).[18] In fact, liberal realism is just an elaboration on solid metaphysical realism since the latter merely prohibits a *general* conflation of metaphysics with epistemology. In contrast, the former explicitly states[19] in which specific cases this general policy might be overruled.[20] We can put the same matter in different terms: Liberally conservative realism (or simply liberal realism) is *liberal* because it is open to a mutual influence of epistemology and meta-

[18] The requirement to keep metaphysical and epistemic considerations generally separated follows, to my mind, from these commitments to (solid) metaphysical realism.

[19] Liberally conservative metaphysical realism at least would, if it was a worked out position, explicitly state under which exact conditions metaphysical conclusions can be drawn from epistemic considerations and thereby indicate in which circumstances solid metaphysical realism admits exceptions from its general proscription to keep metaphysical and epistemic considerations separate. Unfortunately, liberally conservative metaphysical realism is, at the moment, merely a programmatic idea that there are cases – e.g., if the metaphysical phenomenon to be investigated is mind-dependent – where the boundaries between distinct philosophical disciplines can legitimately be transgressed; at least if the specific transgression in question can be backed up argumentatively in a way which does not beg the question regarding the requirement of keeping metaphysical and epistemic investigations apart in general.

In order to provide at least a tentative example to illustrate where else the borders between metaphysics and epistemology tend to become porous, we can point to qualitative aspects of certain mental phenomena where the qualitative aspects coincide with the defining conditions of the phenomena in question: for example, the painfulness of pain. That we cannot fail to recognize when we are in pain, because having pain and feeling it amount to the very same thing, indicates that definition and criteria coincide in this case. This suggests, I think, that also regarding the topic of pain it would be legitimate to draw certain metaphysical consequences on the basis of epistemic considerations – and maybe also *vice versa*. After all, pain is a topic which exhibits quite exotic features, philosophically speaking, since the case where criteria for recognition coincide with metaphysical definition/essence of the phenomenon in question is certainly not the default case. (We might also frame this extraordinary case in a more transcendentalist fashion: In rare cases, where the conditions of the possibility of experiencing a phenomenon coincide with the conditions of the possibility of the phenomenon itself, the boundaries between metaphysics and epistemology might be legitimately transgressed.) In this sense, pain as well as reasoning are rather exceptional and remarkable cases, and they are quite different from the majority of phenomena we strive to explain. I therefore think that we should be rather cautious in regards to where we permit an interplay between epistemology and metaphysics – or philosophy of mind, to be more precise, in the case of pain and in the case of reasoning.

[20] However, for two good examples of how not to draw metaphysical conclusions from epistemic considerations, see Young (2018, § 2.2).

physics, and it is *conservative* [21] because, without a convincing argument to the contrary, it will keep the strict segregation between the domains of metaphysics and epistemology intact. In case of doubt, we may say, any mixing of epistemic with metaphysical considerations is to be avoided. Only if it can be demonstrated for a particular case that the division between metaphysics and epistemology can legitimately be bridged, the interplay will be admitted for the case in question. This is the conservative part of liberal realism where the benefit of the doubt always redounds to strict solid metaphysical realism in order to *conserve* the boundaries between separate philosophical domains. The liberal part is the readiness to admit and thereby *tolerate* exceptions to the strict segregation between metaphysics and epistemology at all.

8.4 The Status of the Principle of Expressibility, Part 3: Analyticity Again

We left our discussion of the question whether the Principle of Expressibility is an analytic truth in section 7.3 with the meager outcome that Navarro-Reyes's (2009) suggestion to explain the analytic status of the principle fails. This does not necessarily mean that the Principle of Expressibility is not an analytic truth. For the Principle of Expressibility to be analytically true means that it is true in virtue of its meaning. Since the principle claims that there is no meaning which cannot be expressed *in language*, its truth – given that it is an analytic principle – depends on the meaning of "language" and the question whether the definition of "language" warrants the claim that every possible meaning can be expressed in language. This certainly does not follow *trivially* from any plausible definition of what "language" means, but the list of characterizing features of language provided in section 2.3 may help to save the Principle of Expressibility as an analytic principle.

The feature of openness, which is plausibly present in every natural language, can help here. Openness is probably not an essential or necessary feature of any language. Formal languages, for example, arguably do not

[21] We might also simply say that it is *realism* for the very same reason.

exhibit the feature of openness – notwithstanding the fact that they are languages nevertheless. Yet it is plausible to assume that every natural language is prone to growth and development in accordance with openness. This means that any natural language, in case it might lack expressibility in a certain area, can always be augmented with (additional) features to compensate for a given limitation. Any language which exhibits the feature of openness therefore has a good claim to satisfy the Principle of Expressibility. Should there be a meaning which cannot be properly expressed in a given language yet, the language in question can always be amended to provide an adequate expression for the hitherto inexpressible content. So, the amount of credibility natural languages – given their openness – lend to the Principle of Expressibility should not be underestimated.

The possible limitations in expressibility which can be overcome due to openness are substantial limitations of languages themselves, not limitations of their speakers. That a certain content does not seem expressible to a speaker in a given language due to a lack of proficiency of said speaker is not a limitation of the language and cannot be compensated with a language's openness to enhancement. Only a genuine incapacity of the language itself to express an intended content needs to be overcome by exploitation of openness of a language. A remedy to overcome restrictions in a language's expressibility may include semantic extension (i.e., expansion of the lexicon) as well as structural (i.e., grammatical) modification of the language in question. Semantic extension might range from trivial cases, such as adopting a loanword which expresses the wanted meaning, to more creative and complex cases like finding a way to figuratively convey the content to be expressed. (For a discussion of the Principle of Expressibility in connection with figurative language, see section 9.1.)

So, can the Principle of Expressibility rightly be called an *analytic* claim? The question is not easy to decide and will probably not permit an uncontroversial answer since analyticity is not an uncontroversial concept.[22] Following a traditional understanding of what "analytic" means, a sentence is analytically true (or false) *iff* whether it is true (or false) depends on its meaning

[22] At least since Quine's (1951b) famous attack on the notion of analyticity.

alone, i.e., *iff* its truth value is independent of empirical facts – except the arguably empirical or contingent facts regarding what certain words mean. Given a plausible explication of the term "language" (as presented in section 2.3), openness is a feature which many languages – probably even all paradigmatic examples of language, especially natural languages – exhibit. If we evaluate the Principle of Expressibility on the basis of this notion of a language (i.e., as including openness), we might plausibly say that the Principle of Expressibility is analytically true.[23]

On the other hand, that most paradigmatic languages exhibit the feature of openness seems to be an empirical and therefore entirely contingent fact which can hardly count as being included in the very meaning of the word "language." Even if we found out that openness is an essential feature of language, this would probably amount to a necessary truth a posteriori. So the Principle of Expressibility is certainly not true by definition alone – as per another traditional formulation of what analyticity amounts to – since openness is not part of the *definition* of "language." Even if openness turned out to be part of the nature of language, and would therefore belong to the essence or *metaphysical* definition of language, it would still not count as being part of the *nominal* definition of "language" (at least not uncontroversially). In any event, the Principle of Expressibility could plausibly count as being analytically true only if openness were part of the nominal definition of language.[24]

[23] Kannetzky's (cf. 2001, p. 200) notion of an 'open language' coincides, I think, with what I call a language which exhibits openness. That Kannetzky (cf. 2001, p. 200) excludes Esperanto from the set of open languages and sorts it with 'fixed languages' (e.g., formal languages) instead, which "[...] are limited in their vocabulary and their syntax" (Kannetzky, 2001, p. 200), stems from a misunderstanding of what (kind of language) Esperanto is, rather than from a disagreement in our notions of open(ness of) language. For ample evidence that Esperanto needs to go with natural languages and not with formal languages when we wish to sort out open and fixed languages, see section 5.2. See also Wells (2009, p. 375), quoted on page 119.

[24] I argued in section 2.3 that openness is not a necessary feature of language since there are languages which do not exhibit openness. What Kannetzky (2001, p. 200) calls 'fixed languages' would plausibly fall in this category of languages which do not exhibit the feature of openness, e.g., formal languages, programming languages, and probably bee language. I do not, however, insist on the claim that there are languages which do not exhibit the feature of openness. I merely think that my characterization of language from section 2.3 is less controversial if openness is not included in the list of necessary features

So, the legitimacy of a claim to the effect that the Principle of Expressibility is analytic(ally true) on the basis of these considerations is quite dubious, to say the least. The problem with settling the question whether the Principle of Expressibility is analytic seems to stem from a certain unclarity whether openness is part of the meaning or definition of "language." These questions need to be settled before we can qualify the Principle of Expressibility as analytic in a traditional understanding of the term – i.e., being true (or false) by definition or in virtue of meaning alone. We might be able to circumvent the problems we encountered in this regard by taking recourse to a slightly different notion of analyticity.

8.4.1 Kripkean Analyticity

Another approach to understanding analyticity shall be mentioned at this point since it might help to shed some additional light on the question of whether we should count the Principle of Expressibility among the analytic claims (or even among the analytic truths).[25] Saul Kripke (1981) famously argued for a clear distinction between the notions of analyticity, apriority, and necessity – and correspondingly of course also for a clear distinction between the notions of syntheticity, aposteriority, and contingency (cf. Kripke, 1981, pp. 34-39). According to Kripke, analyticity/syntheticity is a category of semantics (concerned with the meaning of sentences), apriority/aposteriority is a category of epistemology (concerned with how claims can be justified), and necessity/contingency is a category of metaphysics (concerned with whether

of language. In case it might turn out that demarcating languages from non-languages along the lines of openness is advantageous, I am ready to include openness among the necessary features of language. I also leave the question open whether openness might be a sufficient condition to count as a language.

[25] Note that I use the expressions "analytic" and "analytical" (as applied to sentences, claims, judgments, etc. – but not as applied to philosophy, of course) as meaning "either analytically true *or* analytically false." Analogous stipulations hold for "necessary" (meaning "either necessarily true *or* necessarily false"), "contingent" (meaning "either contingently true *or* contingently false"), "a priori" and "a posteriori" (meaning "either true a priori/a posteriori *or* false a priori/a posteriori"), and "synthetic" (meaning "synthetically true *or* synthetically false").

things could have been different from how they in fact are).[26]

Not only did Kripke provide a clear demarcation of the notions and domains in question, he also insisted that any dependence of one of these categories on any other amounts to a substantial philosophical claim which requires a solid argument and must certainly not simply be presupposed – as has frequently happened in the history of philosophy (cf. Kripke, 1981, pp. 36-39).[27] So, put in different terms, the general outlook Kripke suggests is that we cannot simply derive a statement's[28] metaphysical status from its epistemic status or *vice versa*. The epistemic and metaphysical status (which are the categories Kripke focuses on) of a statement are independent of each other, and the corresponding domains consequentially need to be kept separate.

Nevertheless, Kripke eventually provides a reductive definition of analyticity. According to Kripke, a sentence is analytic *iff* it is necessary *and* a priori (cf. Kripke, 1981, p. 39; p. 56, n. 21; and p. 122, n. 63). Correspondingly, every sentence which is either contingent or justifiable only a posteriori comes out as a synthetic sentence. Kripke's definition can be called a *reductive* definition because the question of whether a sentence is analytic or not (i.e., its semantic status) depends entirely on the questions of whether the sentence in question is necessary (i.e., its metaphysical status) and whether it can be justified a priori (i.e., its epistemic status). The semantic status of

[26] I discussed Kripke's arguments for a strict distinction between metaphysics and epistemology in more detail (and in German language) in my (2014, pp. 65-79).

[27] For a pointed summary of the historical situation and the impact of Kripke's considerations, see Putnam (1975, p. 151).

[28] I will, more or less interchangeably, talk of a *statement's* or a *sentence's* metaphysical, epistemic, and semantic status as the default options for ease of exposition. But the considerations covered here should also be applicable if the entities which have a metaphysical, epistemic, and semantic status should turn out to be propositions, believes, judgments, or any other kind of things. It might also be the case that the epistemic status needs to be attributed to a different entity from the entity which has the metaphysical status. Keith Donnellan (2012, pp. 202 f, n. 2), for example, suggests that the metaphysical status needs to be attributed to a proposition, while the epistemic status is properly attributed to a sentence which expresses the proposition in question. A virtually identical suggestion can also be found in Kaplan (1989, p. 539). I ignore subtleties which might result from considerations of this sort, because I think that they do not change anything regarding the central point in question here – namely that metaphysics and epistemology are to be considered as independent domains.

a sentence is therefore no longer an independent category but is *reduced* to its metaphysical and epistemic status.

It is to be expected that not every scholar will be satisfied with this definition of "analyticity." That said, I suggest putting aside for the moment qualms which may arise regarding this definition in particular or any reductive definition of analyticity in general. Instead, let us see whether this sharpened understanding of analyticity (be it correct or not) may prove fruitful in the current context, namely regarding the question of whether the Principle of Expressibility should be considered to be analytic.

8.4.2 The Metaphysical Status of the Principle of Expressibility

So, how does the Principle of Expressibility fare given Kripke's definition of analyticity? In order to count as analytically true in Kripke's sense, the principle needs to come out as being necessarily true and justifiable a priori. Let us consider these characteristics in turn and start with necessity: What can be said about the metaphysical status of the Principle of Expressibility? To answer this question, we might ask (building on the considerations presented on pp. 208 ff) whether openness is a contingent feature of language, i.e., the question whether language could have lacked this feature. In order to answer this question, we plausibly need to clearly distinguish between the question

(l'') whether any *particular* language with this feature could have failed to exhibit openness

on the one hand and the question

(L'') whether language *per se* – or language in general – could have not had the feature of openness

on the other hand.

Since, by preliminarily adopting his definition of analyticity, we are momentarily operating in a Kripkean framework in any case, we should also

make use of possible world semantics to clear up the issue at hand. As a result of rephrasing the questions mentioned before in possible world semantics terminology, we end up with the following formulations:

(l) Is there a possible world where a given language which exhibits openness in the actual world lacks this feature?

and

(L) Is there a possible world where language *per se* (or language in general) lacks the feature of openness?

We might mirror these questions in modal logic with the following formulas:

$$(l') \qquad \exists x\Big(\big(Lx \wedge Ox\big) \wedge \Diamond\big(Lx \wedge \neg Ox\big)\Big)$$

and

$$(L') \qquad \Diamond\neg Ol$$

respectively.

The first formula (l') says that there is an x, which is a language (Lx) and exhibits openness (Ox) in the actual world, and that there is a possible world where this same x is a language as well but does not exhibit openness. If this formula comes out true, then the first question (l) is to be answered in the positive. The second formula (L') merely claims regarding language in general, represented by the singular term "l", that there is a possible world where l does not exhibit openness (O). If this formula comes out true, the second question (L) has a positive answer.

The question whether openness is a contingent feature of language in the understanding of (l) was only introduced in order to set it aside. The openness (and, by the same token, the possibility to extend a language's expressive capacities) of any particular language is utterly irrelevant for the Principle of Expressibility. The principle merely claims that nothing is in principle and generally inexpressible in language, but it does not make any claim about

a particular language, as already stated at the beginning of section 7.2 on page 173. It is, for example, not necessary – according to the Principle of Expressibility – that every meaning needs to be expressible in one language. The principle is compatible with the fact that no particular language has the capacities to express every possible meaning. For the Principle of Expressibility to be true, it is sufficient that all languages taken together (including merely possible languages and including merely possible extensions of existing or possible languages) provide the resources to, at least in principle, express every possible meaning. This condition is captured in (L), but not in (l). We can therefore set (l) and (l') aside and focus exclusively on (L) and (L').

Although (L') looks much simpler than (l'), the corresponding question (L) might not be easily answerable because the ontological status of language in general is rather unclear. Under which circumstances can we attribute a property to language in general? Is language in general simply an abstraction from all particular languages in the sense of the least exclusive common denominator, i.e., a collection of only those properties and features which every (possible) language has? If so, we might decide to attribute a feature to language in general just in case every particular language has the feature in question, so that nothing could be a language which lacks any of these features. Or is language in general an independent object – perhaps something like the Platonic form of language? If so, how do we find out about language in general, given that – pace Russell (1911) – we might not be acquainted with Platonic forms or other universals (cf. Benacerraf, 1973)?

I suggest that we treat language *per se* as the most inclusive abstraction from all possible particular languages, comprising every feature which any particular language could possibly have. This suggestion will leave us with language in general as a quite strange entity: being at the same time possibly inflectional[29] and possibly isolating,[30] possibly tensed[31] and possibly not

[29] Like, e.g., Latin and German, together with the majority of the stereotypically European (Romance) languages.

[30] Like, e.g., Vietnamese, Mandarin, and – to a lesser degree – also (contemporary) English. (Old English used to be much more inflected than Modern English.)

[31] Like arguably all natural languages are, even Hopi – pace Whorf (1956).

exhibiting any tense structure at all,[32] at the same time, possibly, organically developed over several generations with a rich cultural background,[33] as well as possibly invented from scratch,[34] etc. Language *per se* will therefore have seemingly incompatible features, such as possibly being agglutinative and possibly not being agglutinative.[35] That language in general combines all these features makes it, importantly, only *seemingly* inconsistent since all features – including mutually exclusive features – are attributed to language in general merely in the "subjunctive mood," so to say.[36]

Language *per se* in this sense might be a strange entity, but I take talk about language *per se* in this sense to be no stranger or less permissible than talk about *the mathematical set* in the abstract, which might be either empty or might have infinitely many members, or which might even contain itself or might not. I think that traditional critiques against such abstract(ed) entities – famously to be found in Berkeley (1999b, §§ 6-16, pp. 9-17), for example – are unproblematic as long as we do not expect that we might come across such abstract entities in the "real world."[37] Assuming such abstract entities is, I claim, fruitful and warranted for theoretical purposes; it is not more problematic than talk of the common average family with 1,47 children.[38] So, I claim, language in general is also a legitimate theoretical notion as long as we are not fooled into looking for a language course where we could learn to speak language *per se*.

In order to make this notion of language *per se* more tangible, we may think of it in analogy with Universal Grammar, famously advocated by Noam

[32] Like, for example, formal languages of propositional or predicate logic or generally programming languages.

[33] Like national languages generally are.

[34] Like several constructed languages, e.g., the examples mentioned on page 99, including Esperanto and other conlangs. Note, however, that the property of being invented from scratch does not prevent a language from having a rich cultural background. (See section 5.2.)

[35] For an explanation of the term "agglutinative," see pp. 116 ff.

[36] For a clarification in more formal terms, see pp. 166 f.

[37] As far as Berkeley is concerned, as long as we do not confuse the idea (as a psychologically occurring entity, which might need to be a concrete particular) with what it represents, i.e., its content (which can nevertheless be abstract). (Cf. Berkeley, 1999a, p. 210)

[38] This number is made up and not meant to represent actual statistical data.

Chomsky (among others). By hypothesis, we can generate the grammatical structure of every human language from Universal Grammar by setting parameters which are determined by the Language Acquisition Device (or language faculty). In this way, i.e., by setting more and more parameters, we can restrict the options "contained" in Universal Grammar until we arrive at a structure which describes actually spoken languages. In this sense also Universal Grammar – in a way – comprises every possible (grammatical) feature of (human) language, presumably even features which cannot appear together in any single human language. So, universal grammar also plausibly has incompatible features *in potentia* and marks the limits of possible (human) languages. Similarly, what is not contained in language *per se* is not a possible (human or non-human) language at all.[39]

In any case, given the definition of language *per se* as including every possible feature of every possible particular language, the answer to question (L) is a clear "No." Language in general, understood in the sense just mentioned, will have all possible features of every possible language in every possible world. Mind that I presuppose system S5, where there are no restrictions on accessibility between possible worlds, since the accessibility relation in S5 is an equivalence relation – meaning a reflexive, symmetrical, and transitive relation – so that every possible world is accessible from any other possible world. It trivially follows that, since language in general has every possible feature of every possible language in every possible world, language in general also has the feature of openness in every possible world.[40] It is therefore necessary that language *per se* has the feature of openness, so

[39] Note that in order to make use of the notion of Universal Grammar to shed light on the question of what language *per se* might "look" like, we do not need to presuppose that Universal Grammar is real (in whatever meaning of this term), innate, genetically endowed, or any other potentially controversial claim about Universal Grammar. The analogy can do its work as long as we understand how Universal Grammar can comprise features which are mutually exclusive in fully specified languages. No substantial commitments regarding Universal Grammar should be taken to follow from the explicative analogy with language *per se* or language in general.

[40] At least in every possible world where language in general exists. I leave it open whether language *per se* is a necessary entity which exists in every possible world. Following Kripke (cf. 1981, p. 48), I take it to be sufficient to say that a feature F is a necessary feature of x if x has F in every possible world where x exists.

that the negation of (L') comes out true.[41]

We have gone halfway towards answering the question of whether the Principle of Expressibility can count as analytically true in the Kripkean sense. Since (L') is false and the answer to question (L) is "no," we can plausibly conclude that the Principle of Expressibility has its truth value necessarily. That is, it is necessarily true if it is true (and necessarily false if it is false), given that openness is a necessary feature of language *per se* and that openness guarantees that the Principle of Expressibility is true. I do not consider the Principle of Expressibility to be proven yet. I only take it to be settled at the moment that the Principle of Expressibility has its truth value necessarily – whichever it may be – and so the Principle of Expressibility is still in the running for the status of being an analytic principle. In order to show that it is an analytic principle, we need to show that the Principle of Expressibility is also a priori, in addition to its being necessary, in the understanding of the principle which results from connecting it with question (L).

8.4.3 The Epistemic Status of the Principle of Expressibility

In order to settle whether the Principle of Expressibility can be known (to be true or false) a priori – in the same understanding of the principle, i.e., building on question (L) – we need to ask whether we can justify an answer to question (L) a priori, i.e, without recourse to empirical knowledge. Whether this can be done will, of course, depend on what counts as empirical knowledge, and this might not be so easily decidable.

[41] "Language in general necessarily has the feature of openness" (or, equivalently, "It is necessary that language *per se* has the feature of openness") will be formalized as $\Box Ol$, which is equivalent with $\neg\Diamond\neg Ol$, and that is just the negation of (L') which has it that $\Diamond\neg Ol$. That every individual language exists only contingently is irrelevant for this claim. I take language in general to exist at least in every possible world where any language exists. There are also possible worlds where no languages exist and where plausibly also language *per se* does not exist. As noted in the previous footnote (40) already, language *per se* still has the feature openness necessarily, so that there is no possible world where language in general exists and lacks the feature of openness.

Can we only empirically know that language *per se* has the feature of openness because we can only know empirically that some languages (among those are some actually existing languages but also several merely possible languages) exhibit this feature? If so, I tend to side with Kripke by claiming that the conclusion that language in general has the feature of openness "[...] is known a posteriori, since one of the premises on which it is based [– namely that there are languages (possible or actual) which exhibit the feature of openness –] is a posteriori." (Kripke, 2011a, p. 17) Since the answer to (L) is known a posteriori and since the answer to (L) is a crucial step in determining the status of the Principle of Expressibility, the epistemic status of the answer to (L) might be inherited by the Principle of Expressibility, which would make it an a posteriori truth (or falsity).

All the same, that we can answer (L) a posteriori – by coming to know that there are languages which exhibit openness on empirical grounds – does not automatically mean that we can answer (L) *only* a posteriori. If we can also find an a priori answer to (L) – i.e., if we can justify an answer to (L) without recourse to empirical facts, e.g., without "looking" whether there are actually languages which exhibit the feature of openness – then the Principle of Expressibility might still count as an a priori claim.

Frankly speaking, I do not know whether all of this can be known a priori, only a posteriori, or perhaps even only a priori. It might be that I simply have not sufficiently considered how an answer to (L), and building on it also the Principle of Expressibility itself, can be justified, i.e., only a posteriori, only a priori, or perhaps in both ways. Moreover, perhaps some might say that I am simply confused about the epistemic status of the Principle of Expressibility and that this is the only reason which prevents me from recognizing the Principle of Expressibility as an analytic claim. That might be the case. Nevertheless, as of yet I have not seen a convincing argument for the claim that the Principle of Expressibility is an analytic truth, nor have I seen (or provided on my own) a convincing argument for the claim that the Principle of Expressibility is not an analytic truth. In fact, I do not deem the question whether the Principle of Expressibility is *analytically* true or not of ultimate importance. I primarily want to know whether the

Principle of Expressibility is (plausibly to be held) true or not. Whether it is analytically true, and therefore necessarily true as well as justifiable a priori, is at best of secondary interest.

It might be possible to determine that openness is a possible feature of language a priori. I am sure that we know that openness is a possible feature of language since we know that there are languages which exhibit this feature. I am satisfied with the fact that we know that openness is a possible feature of language and that this knowledge is suited to lend immense credibility to the Principle of Expressibility. I will, given this situation, tolerate that I do not know how exactly we can know about the openness of language (*per se*) – whether only empirically or perhaps also a priori.

I will leave the issue regarding the precise epistemic status of the Principle of Expressibility – and consequently also regarding its precise semantic status, according to Kripke's reductive definition of analyticity – without a conclusive answer. I will only say that besides its unclear epistemic and semantic status, it is hard to escape a certain impasse when we try to settle the truth value of the Principle of Expressibility, which is my main concern here: to find out whether the Principle of Expressibility is true or false, not to find out whether it is analytic or synthetic. On the one hand, we find ourselves in a situation similar to the one we face regarding Goldbach's conjecture (see section 8.3.1). We do not have – at least not currently – a method to prove the Principle of Expressibility. We are therefore not in a position to determine that the principle is true. Yet the situation is, on the other hand, even worse than that regarding Goldbach's conjecture since it seems that we do not even have a method to falsify the Principle of Expressibility. Although Navarro-Reyes's explanation, presented on pp. 183 f, is flawed as a proof of the analytic status of the principle, his claim that we cannot come up with a counterexample to the Principle of Expressibility is nonetheless valid. So, if we have neither a method of verification nor a method of falsification for the principle at hand, how are we to settle its truth value?

I have repeatedly emphasized that the fact that a claim is beyond our epistemic reach does not deprive the claim of its truth value. Any claim

is true or not true,[42] independently of our ability to determine its truth value. So, even if settling the truth value of the Principle of Expressibility was beyond our epistemic means, I presuppose that either it conforms with the facts or it does not. But are we really unable to find out whether the Principle of Expressibility is true or false? Is the Principle of Expressibility an *Ignorabimus*?[43]

I do not think that this is the case. Even if an undeniable demonstration of the truth of the principle seems out of reach at the moment, it is not as if we had no clue whether the Principle of Expressibility might be true or not. That we know that openness is a possible feature of language (and even a necessary feature of language *per se*) – even though we do not exactly know which epistemic status this knowledge has, i.e., whether it is justifiable a priori or not – makes the Principle of Expressibility highly plausible. Given that this provides us with a good reason to believe in the Principle of Expressibility, I think that it might be sufficient to rule out the most pressing (and possibly) remaining doubts regarding the principle and to warrant confidence towards the Principle of Expressibility which at least comes close to saying that we *know* that it is true. This might not give us the confidence we would like to have in a philosophical claim, but we are generally content to say about many things that we know them even though they were never stringently demonstrated. We take many things to be part of our stock of knowledge, and often for good reason, on the basis of the fact that they were made highly plausible. I shall be satisfied if I can achieve that and leave it to someone else to provide an indisputable proof of the Principle of Expressibility.

I intend to achieve said level of plausibility – which allows us to claim that we know the Principle of Expressibility to be true even without a demonstrative proof – by dispelling the, as far as I can see, most worrisome problem cases for the principle. The most worrisome cases are, to my mind, figura-

[42] I do not say "true or false" because I prefer to stay neutral regarding the question of bivalence. However, I am more confident regarding the principle of excluded middle, and I presuppose that any claim we take into consideration here is not so ill-framed that we need to consider it meaningless.

[43] See Emil Du Bois Reymond (1884, p. 46), who famously claimed that there are not only things we do not know – *'Ignoramus'* – but also things we will never know because we cannot know them – *'Ignorabimus'*.

tive language use – notably metaphor – and qualitative aspects of experience, which may seem notoriously elusive from proper (and purely linguistic) formulation and thereby also expression. The following chapter is therefore dedicated to showing that the existence of metaphor and of qualitative phenomenological content do not provide a cogent argument against the Principle of Expressibility. This, I claim, will put us into a position to believe in the Principle of Expressibility with sufficient confidence to say that we know that no content whatsoever is, in principle, ineffable.

Chapter 9

Problem Cases for the Principle of Expressibility

Given the previous discussion of language *per se* and the plausibly important role which the feature of openness probably plays for the Principle of Expressibility, it seems prudent to consolidate the principle's credibility by clearing out what is commonly conceived as the most pressing problem cases for (adequate) linguistic expression, and thereby also for the Principle of Expressibility. As soon as the most prominent apparent counterexamples to the Principle of Expressibility are dispelled, we lack good reason not to believe that the Principle of Expressibility is true. Given the already established plausibility of the Principle of Expressibility, I claim that we then have, by the same token, good reason to believe in the Principle of Expressibility. In order to get there, this chapter discusses the two most plausible problem cases for the Principle of Expressibility in turn: first (in section 9.1) metaphor and then (in section 9.2) phenomenal character.

9.1 Metaphor and the Principle of Expressibility

9.1.1 Stating the Problem

It might be doubted that certain forms of language use still conform with the Principle of Expressibility. Cases which come to mind will certainly include metaphorical use of language, as well as making use of analogies to convey a given idea. In general, any kind of figurative or indirect way of expression might be problematic for the Principle of Expressibility. Do these kinds of language use still comply with the Principle of Expressibility, given that it demands that it is possible to find an exact or explicit expression for any given meaning? Why do we frequently use indirect or metaphorical ways of expression at all if we could – according to the Principle of Expressibility – say what we want to say directly and straightforwardly as well? Even more importantly, do we not at least sometimes feel forced to make use of figurative or indirect speech in order to express what we wish to say? Would it not be extremely problematic for the Principle of Expressibility if these ways of communication turned out to be indispensable? These matters remain to be seen.

The relationship between the Principle of Expressibility and metaphor has traditionally been rather tense. John Searle emphatically claims that it needs to be possible to paraphrase any metaphorical expression without any residue of surplus meaning on the side of the metaphor. In other words, Searle claims that every metaphor can be entirely paraphrased away. If it is not possible to find a literal equivalent for every metaphorical expression, Searle thought, the Principle of Expressibility is in danger. This might not be obvious from the start when we take a look at what Searle has to say about metaphors. He concedes, for example, that often, when we try to paraphrase a metaphor to express what it means literally "[...] we feel that the paraphrase is somehow inadequate, that something is lost." (Searle, 1979b, p. 82) Even more explicitly, he says that

> [s]ometimes we feel that we know exactly what the metaphor

> means and yet would not be able to formulate a literal [paraphrase] sentence because there are no literal expressions that convey what it means. (Searle, 1979b, p. 83)

This does not sound as if Searle thought that it must be possible to paraphrase metaphors away if the Principle of Expressibility is true. However, concluding that Searle takes a liberal approach towards potentially unparaphrasable metaphors would be premature. So let us hear him out.

In his own theory about metaphors, presented in Searle (1979b), he presupposes right away that metaphorical utterances can be paraphrased without any problem:

> Because in metaphorical utterances what the speaker means differs from what he says (in one sense of "say"), in general we shall need two sentences for our examples of metaphor – first the sentence uttered metaphorically, and second a sentence that expresses literally what the speaker means when he utters the first sentence and means it metaphorically. (Searle, 1979b, pp. 81 f)

Hence, the possibility to express what a metaphor means in a literal way is an integral precondition of Searle's theory of metaphor. In the last section of his article about metaphors, Searle reinforces this presupposition:

> The question of whether all metaphorical utterances can be given a literal paraphrase is one that must have a trivial answer. Interpreted one way, the answer is trivially yes; interpreted another way, it is trivially no. If we interpret the question as, "Is it possible to find or to invent an expression that will exactly express the intended metaphorical meaning [...]?" the answer to that question must surely be yes. It follows trivially from the Principle of Expressibility (see Searle, [2011]) that any meaning whatever can be given an exact expression in the language.
>
> If the question is interpreted as meaning, "Does every existing language provide us exact devices for expressing literally whatever we wish to express in any given metaphor?" then the answer is

> obviously no. It is often the case that we use metaphor precisely because there is no literal expression that expresses exactly what we mean. (Searle, 1979b, p. 114)

Searle does not even consider the option that there might be content which – even in principle and in any language – cannot be literally stated but only adequately expressed in metaphor. He takes it for granted that it is always possible to invent new expressions to literally convey meanings which can currently only be expressed metaphorically.[1] Likewise, Searle takes it for granted that we do not always have a literal expression at hand to paraphrase a metaphor away. So, metaphors serve an important practical purpose for language use in several respects: They allow us to express content we have no literal way of expressing (yet) and therefore "[. . .] serve to plug [. . .] semantic gaps [. . .]" (Searle, 1979b, p. 83), and they can also help us achieve artistic or poetic effects which would be lost in literal communication. Nonetheless, metaphors are, according to Searle, not theoretically indispensable because every metaphor can – in principle – always be paraphrased away.

That Searle claims that "[i]t follows trivially from the Principle of Expressibility [. . .]" (Searle, 1979b, p. 114) that "[. . .] all metaphorical utterances *can* be given a literal paraphrase [. . .]" (Searle, 1979b, p. 114; emphasis added), even if there is none available at the moment, clearly shows that he conceived of the Principle of Expressibility as being in conflict with the possibility of unparaphrasable metaphor. However, I think that Searle was mistaken in this regard. The position that metaphorical expression cannot always be avoided – let alone that it should always be avoided at any cost – does not stand in any conflict with the Principle of Expressibility whatsoever.

9.1.2 How to Conceive of Metaphor

To flesh out this idea, let us take recourse to a radically different view on metaphor from Searle's. Although Searle (1979b, p. 77) is very explicit about

[1] This is, by the way, additional evidence that the modality in Searle's formalized expression of the Principle of Expressibility should be taken to mean "feasibility" rather than "possibility" in the sense of possible world semantics. See pp. 166 and 168 in section 7.1.

the fact that "[m]etaphorical meaning is always speaker's utterance meaning" and that it is crucial to distinguish "[...] between sentence or word meaning, which is never metaphorical, and speaker or utterance meaning, which can be metaphorical" (Searle, 1979b, p. 86), metaphor is for him, it seems, still basically a matter of language and expression. Not so for George Lakoff (1993), who claims "[...] that the locus of metaphor is thought, not language, [because] metaphor is a major and indispensable part of our ordinary, conventional way of conceptualizing the world [...]" (Lakoff, 1993, p. 204). According to Lakoff, we do not only – more or less occasionally – speak in metaphors, but we generally *think* in metaphors since "[...] as soon as one gets away from concrete physical experience and starts talking about abstractions or emotions, metaphorical *understanding* is the norm." (Lakoff, 1993, p. 205; emphasis added) The very conceptual systems we use to understand ourselves and the world around us are metaphorical, so "[t]he metaphor is not just a matter of language, but of thought and reason." (Lakoff, 1993, p. 208) It can even be shown that, Lakoff (1993, p. 222) claims,

> [...] the most common abstract concepts – TIME, STATE, CHANGE, CAUSATION, ACTION, PURPOSE and MEANS – are conceptualized via metaphor. Since such concepts are at the very center of our conceptual systems, the fact that they are conceptualized metaphorically shows that metaphor is central to ordinary abstract thought.

Lakoff (1993, p. 244) summarizes his main findings about the nature of metaphor as follows:

> Metaphor is the main mechanism through which we comprehend abstract concepts and perform abstract reasoning.
>
> Much subject matter, from the most mundane to the most abstruse scientific theories, can only be comprehended via metaphor.
>
> Metaphor is fundamentally conceptual, not linguistic, in nature.
>
> Metaphorical language is a surface manifestation of conceptual metaphor. [...]

I will not set out to discuss whether Lakoff's conceptual theory of metaphor is correct or not. I merely introduce it as a canvas to motivate the option that Searle's claim that every metaphor can be paraphrased away may be mistaken. If we assume that speaker's meaning might be metaphorical because we think metaphorically, then it seems plausible to assume that metaphorical meaning can only be expressed exactly and adequately in a metaphorical way. Any literal and therefore not metaphorical paraphrase must fall short of an exact and adequate expression of the intended meaning because the intended meaning was metaphorical in the first place.[2]

In order to avoid a terminological confusion looming here, we should not say that the exact expression of a genuinely metaphorical thought should not be called "metaphorical" anymore. We might think that in such a case we are actually confronted with a *literal* expression of a metaphorical content. This, I think, is just a confused way of looking at things. An expression is not metaphorical because it does not literally, but merely metaphorically,

[2] I presuppose at this point that an inherently or genuinely metaphorical content (given that something like this exists) can only be accurately rendered in a metaphorical expression. This potentially controversial assumption is necessary because it cannot be generally presupposed that expression and content need to share each and every feature if the expression is to count as an accurate expression of the content in question. Many properties cannot even be possibly shared between content and its expression. The property of being printed in boldface, for example, can only be exemplified by an (instantiation of an) expression, strictly speaking. To say of the content that it is printed in boldface can only be rather loose talk or a gross category mistake. A requirement to the effect that content and expression need to share, e.g., the property of being mental in order for the expression to count as exactly presenting its (mental) content would probably render each and every attempt to accurately communicate anything impossible from the outset. Therefore, being (inherently or genuinely) metaphorical is, I hold, a quite special property if it needs to be shared among content and expression in order for the expression to count as accurately expressing the intended (metaphorical) content.

I do not have, at the moment, a convincing argument to exclude the possibility that there might be a sentence whose literal meaning exactly matches an intended metaphorical content. (I can merely point out that the very purpose of calling a content "genuinely metaphorical" is to suggest that it cannot be expressed literally. All the same, a mere stipulation like this falls, of course, considerably short of an actual argument.) I am having a hard time imagining such a case, but it might be possible – for lack of proof of the opposite. However, if (genuinely) metaphorical content could even be expressed literally – which is congenial with what Searle claims – then my job of defending the Principle of Expressibility becomes rather easier than harder. So, there is no reason to vigorously argue against this possibility. I wish to thank Frank Hofmann for pointing out the need to elaborate on these points in a draft of this chapter.

express an intended content. An expression is metaphorical because what it is supposed to mean or express differs from what it means when we take the (metaphorical) statement in its literal meaning. If someone tells you that you are their sunshine, the utterance "You are my sunshine" is certainly metaphorical even if it exactly expresses (in its metaphorical meaning) what the speaker wants to say – given that the speaker did not mistake you for a stream of photons. In any case, being an exact expression of an intended meaning is not the same as being a literal expression of an intended meaning. Therefore, the idea to call an utterance which gives exact expression to a metaphorical content a "literal expression of metaphorical content" amounts to a confusion of exactness/adequacy/accuracy with literalness.[3]

Be that as it may, that we think in metaphors is, again, not to be defended here as the correct theory about our conceptual mental life. Whether the conceptual theory of metaphor is correct will be left open here. What is relevant at this point is that a theory such as Lakoff's can motivate the claim that metaphorical ways of expression are, pace Searle, indispensable. This is because a metaphorical content can plausibly only be adequately and exactly expressed by means of a metaphorical utterance. Is this option in conflict with the Principle of Expressibility as Searle insinuates?

I claim that it is not. If there should be content which cannot be (adequately) expressed in a literal way because the content in question is inherently metaphorical, such meaning can still be accurately expressed by means of a metaphor. Since metaphorical ways of expression are a part of many languages' expressive capabilities, nothing should block us – in principle – from finding adequate expressions for metaphorical content. So, even if not

[3] The question might come up why I hardly talk about explicitness anymore in connection with the Principle of Expressibility: Explicitness is the central notion in how Searle makes use of the Principle of Expressibility. The use I make of the Principle of Expressibility considerably transgresses the "merely" methodological use of the principle found in Searle's (2011) discussion about *Speech Acts*. In discussing the relation between what can be said and what can be thought in the present context, it is not sufficient that every speech act can be conducted by using an *explicit* expression. I need to argue, in addition, that every possible (mental) content can be *accurately* expressed. The expressions "adequate" and "exact" (and grammatical variants thereof) are sometimes used as variations for what "accurate" (and its grammatical variants) means in the sense specified by Binkley (1979) and discussed here in section 6.2.6.

every meaning can be exactly paraphrased in literal language, language still provides the means to exactly express a given content metaphorically. Consequentially, genuinely metaphorical content is no threat for the Principle of Expressibility since, from the possibility that not every content can be expressed literally, it does not follow that there might be content which cannot be expressed exactly, accurately, or adequately.

9.1.3 *Utterance Meaning* vs. Speaker's Meaning vs. Sentence Meaning vs. "Hearer's Meaning"

This is the right time to introduce *utterance meaning*, in addition to and in order to distinguish it from speaker's meaning and sentence meaning, as already announced on pp. 137 f. Identifying what can be said, i.e., what is sorted into set S, with sentence meaning – as I have done up until now – is good enough for most parts of the debate, but it is not entirely accurate. Drawing on the insight that "[...] we can do quite different things using the same words$_{[,]}$" (Kannetzky, 2001, p. 193) we need to acknowledge that the actual message conveyed in a given utterance is most often dramatically underdetermined by the literal meaning of the words uttered, i.e., by sentence meaning.[4]

> Searle's example "The cat is on the mat" reminds us that we normally assume that the cat sits or lies on the mat. However, the sentence itself does not contain this information and it does not follow logically. The sentence could also say, for example, that the mat stands vertical and the stiff-made cat is glued on the upper edge. (Kannetzky, 2001, p. 204)

Now, does that mean that the sentence "The cat is on the mat" is insufficient to exactly – i.e., properly and explicitly – express the thought that the cat is on the mat (in a usual, unsurprising way)? Kannetzky (2001, p. 204) goes on to elaborate on the case:

[4] See also pp. 156 f.

> If a speaker means that the cat is on the mat, there must be an exact linguistic expression of it. What the speaker means is without doubt "The cat is on the mat" in the usual, normal sense. How can this normal sense be explicated? Is it necessary to state explicitly that the cat is not glued on the upper edge of a vertical[ly] standing mat? Is it necessary to explicate the whole background of the sentence for capturing its usual meaning? Can this be done by uttering further sentences in order to explain the conditions of the applicability of the first one? This would lead to an infinite regress of explication. No speech act could be explicated – the [P]rinciple of [E]xpressibility would be empty and meaningless, because it would require something impossible. In contrast, we know that we do not need an infinite number of sentences for expressing something exactly. The recourse ends practically after few steps. We tacitly refer to this common background of orientations and practical abilities that do not leave space for persistent doubts. At least we can eliminate such doubts by realizing parts of this background (especially the real or possible uses of the expressions in question).

The pragmatically enriched meaning (not contained in the literal sentence meaning), which Kannetzky talks about, is the utterance meaning. Utterance meaning will in most cases be different from literal sentence meaning. In contrast to literal sentence meaning, which is context invariant, the utterance meaning is saturated with background assumptions to exclude all the logically possible but practically irrelevant contents which could be expressed with one and the same sentence. Every sentence has only one sentence meaning but also the potential to express countless (more or less slightly) different utterance meanings.

Which utterance meaning is expressed in a given speech act by using a given sentence will crucially depend on the context in which the sentence is uttered. If the communicative attempt is successful, the given context will restrict the plethora of possibly intended meanings, which could all be expressed by using the same words, to the correct utterance meaning. Utterance

meaning is, however, not determined by speaker's meaning. Speaker's meaning – i.e., the intended content to be communicated – and utterance meaning can come apart if the communicative attempt is not successful. This might happen, e.g., if the phrasing is rather clumsy or if the speaker makes a slip of the tongue. Utterance meaning is also not determined by what a potential hearer might understand. Every utterance, even if properly expressed, can always be misunderstood. So, utterance meaning is (usually) different from sentence meaning, (might be) different from speaker's meaning, and does not (necessarily) coincide with what a hearer might understand an utterance to mean.

Utterance meaning is therefore an additional dimension of meaning which cannot be reduced to any of the aforementioned meaning dimensions – i.e., speaker's meaning, sentence meaning, and "hearer's meaning" (the meaning which a hearer takes a given utterance to have or the meaning which a speaker attempts to communicate with a given utterance, according to the hearer). Still, utterance meaning can be objectively determined by taking into account the literal meaning of the words uttered, the context of the utterance, and plausible assumptions about common background and expectations of speaker and hearer.[5]

The process of determining utterance meaning is not mysterious at all. It is something we do all the time. But as is often the case with many things we naturally do every day, it is hard to explicate what exactly goes on when we do it or how exactly we are able to do it. In many cases, it will be utterance meaning which we "directly" grasp in a conversation while we would need to make some efforts to envision the literal sentence meaning of what somebody said. Also, it is not sentence meaning but utterance meaning which usually gives us the right clue to what somebody means to say.[6] Furthermore, importantly, a hearer as well as the speaker himself might

[5] If mentioning background assumptions and speaker's and hearer's expectations should strike the reader as being redundant because they are already part of the context, so be it. I do not wish to presuppose any particular theory about what context does or does not include. I do not wish to presuppose, e.g., whether Stalnaker's (2002) 'common ground' is contained in the context. I wish to leave this open and therefore mention background assumptions and expectations separately.

[6] Sentence meaning is usually just a mediating step (taken consciously or unconsciously)

be in error regarding what the true utterance meaning of a given utterance is. In this sense, utterance meaning is objective and independent from both what anybody might understand and what somebody might mean.

Utterance meaning can also be objectively vague or unclear, for example, if the given context is insufficient to exclude conflicting meanings which could plausibly be compatible with the literal meaning of the words uttered: if, for example, there is not enough information available to conclusively resolve anaphoric reference. Think of examples like "Tom and Bill met at the bar, where he invited him for a drink." Who invited whom for a drink? Without clear contextual clues to determine whether "he" refers to Tom and "him" to Bill or the other way round, the utterance meaning is (objectively) unclear. This is so even if the speaker (plausibly, since he is not confused about what he wishes to say) knows exactly who invited whom and even if the hearer gets the reference right by mere luck. Chances are fifty-fifty for the hearer to understand the utterance in the way intended by the speaker, so the odds are not bad. Yet even if there is a clear speaker's meaning and even if what the hearer understands coincides with the speaker's meaning, the utterance meaning is still ambiguous because mutually exclusive salient interpretations are available which are compatible with the literal meaning and all plausibly available contextual clues.

I ignore situations where Bill and Tom are both known to the speaker as well as to the hearer and where it is clear that Bill invited Tom because Bill has a reputation for always inviting everyone and Tom is a famous miser. If this was the actual situation, then background knowledge would plausibly be sufficient to determine utterance meaning and exclude all unintended salient

on the way to detection of utterance meaning – aided by awareness of the context of the utterance, together with background knowledge. Three factors are usually needed to determine utterance meaning: literal sentence meaning, relevant factors regarding the context of utterance, and general (commonly shared) background knowledge. With these three components at hand we can "calculate" utterance meaning, and, if everything goes well, utterance meaning is sufficiently close to speaker's meaning to achieve understanding. If, for example, an utterance of "Well done!" was (clearly) meant ironically, then the utterance meaning will match the intended (ironical) meaning by the speaker whereas taking the literal sentence meaning to be the intended speaker's meaning would lead to a misunderstanding because the irony was missed. Generalizing this point, we can say that utterance meaning will usually include Gricean implicature (cf. Grice, 1991).

meanings. We probably should say that the utterance meaning is clear, in this case, for the speaker and the hearer who share a broad common ground of background knowledge but that the utterance meaning is not clear for the uninitiated bystander who overhears the utterance in question.

We might construct a similar situation by excluding the quite specific and shared background knowledge about Bill and Tom's respective character traits. Instead, we assume that it is a well-known fact for the hearer that the speaker who utters the sentence in question meticulously sticks to one of Grice's maxims of manner – be orderly! (cf. Grice, 1991, p. 27) – and always introduces pronouns in the order of appearance of the names they share their reference with. In this case it would be clear that Tom invited Bill since otherwise the speaker would have mentioned Bill before Tom or he (the speaker) would have put the latter part of the utterance in the passive voice. Also in this case the utterance meaning might be clear for a close friend but probably not for a casual acquaintance of the speaker.

Given these examples it might have been an overstatement to say that the utterance meaning can be objectively determined since it is clear that not everyone will in any situation be in a position to settle what the actual utterance meaning is. Depending on the background knowledge, some people will be able to determine the utterance meaning in a given situation; others will not be in a position to do so. I acknowledge this fact, and I suppose it would be a fair complaint to say that utterance meaning, given the situations just discussed, cannot be called "objective." So be it. What I want to make absolutely clear by calling utterance meaning "objective" is that it is not dependent on what anybody – be it the hearer or even the speaker herself – might *think* an utterance means. So, by calling utterance meaning "objective," I wish to make clear that it is not subjective in the sense of being dependent on somebody's mere opinion, be it speaker, addressed hearer, or bystander.[7] However, if somebody were to insist that this is an insufficient

[7] Probably even introducing an "ideal hearer," who has all the information and is immune to reading meaning into an utterance which was not actually expressed, given the circumstances, would not solve the issue. Such an ideal hearer could probably determine utterance meaning even if it is implausible to assume that any "normal" hearer could do so. In this case, I would prefer to say that the utterance meaning is indeterminate, ambiguous,

reason to call utterance meaning "objective," I am prepared to submit to this objection. The term "intersubjective" might be a better match.

The important point, however, is that in order to have an adequate, exact, and accurate expression for the intended meaning, it is not necessary that speaker's meaning coincides with sentence meaning. This requirement would turn the most mundane communicative situation into an unbearably tedious endeavor because we would be constantly busy trying to exclude even the most remote recherché cases of possible misunderstandings in every formulation. Although it might sometimes feel as if communication in the philosophy seminar room operates this way, this is not how we usually communicate. In many cases, we are effortlessly able to find exact and adequate expressions to communicate what we have in mind. This would be a big mystery if, in order to adequately express ourselves, we needed to find formulations which *literally* mean what we want to say. Yet, this is not the case. To accurately express any intended meaning, we merely need to achieve sameness of content between what is meant and what is said by bringing utterance meaning into agreement with speaker's meaning. This is much simpler than bringing sentence meaning and speaker's meaning into agreement because all sorts of pragmatic factors automatically support us in expressing what we wish to convey with relative ease as long as we stick to common ways of putting things into words. Traditionally established ways of speaking (or writing) provide a huge amount of common ground which allows us to exploit established conventions to our advantage.

I will just give one quick example to illustrate what I have in mind: Common knowledge about what counts as a marked and what counts as an unmarked expression[8] is information about a linguistic feature we can easily

vague, or what have you. I would rather not say that in such a case there is a determinate utterance meaning although nobody can plausibly settle which one it is. Therefore, I refrain from any attempt to settle the problem by introducing an ideal hearer.

[8] Laurence Horn (2006, p. 16) characterizes marked and unmarked expressions as

> [...] two expressions covering the same semantic ground, [where] a relatively unmarked form [is usually] briefer and/or more lexicalized [and] tends to be [...] associated with a particular unmarked, stereotypical meaning, use, or situation, while the use of the periphrastic or less lexicalized [i.e., marked] expression, typically more complex or prolix, tends to be [...] restricted to

make use of to convey intended meaning in a quite economical way. If I say, for example, "I stopped the car," then everyone will probably understand that I did so in the usual way, i.e., by stepping on the brake – given that it is contextually settled that I was the driver of the car in question. If I say, in contrast, "I brought the car to a stop," then everyone will probably understand that I did so in a somewhat unusual way or, at least, that the procedure was more complicated and effortful than it should be – given that it is not lack of language proficiency which accounts for my uncommon way of putting things.

By exploiting features such as the difference between marked and unmarked expressions, which is just part of general linguistic knowledge every competent speaker has at her disposal, I can pack much more information into utterance meaning than literal sentence meaning would ever allow in such an efficient manner. This is also why the utterance meaning of using the sentence "The cat is on the mat" will generally include the information that the cat is lying or sitting on the mat in a usual way. The case in which the cat is glued on top of the upright mat is excluded because such an unusual situation would not be communicated with such an ordinary (unmarked) expression. In this sense, it is not only context which enriches utterance

---

> those situations outside the stereotype, for which the unmarked expression could not have been used appropriately. [Footnote omitted]

He also provides the following example, among others, to illustrate the difference between marked and unmarked expressions: "He got the machine to stop" is a marked form while "He stopped the machine" is an unmarked form (cf. Horn, 2006, p. 16). "The use of the periphrastic causative in [the former example] implicates that the agent achieved the effect in a marked way ([e.g.,] pulling the plug, throwing a shoe into the machine [etc.]) [...]" (Horn, 2006, p. 17). See also Horn (1997, p. 314) where the following example appears: "Black Bart caused the sheriff to die" as a clearly marked expression, and "Black Bart killed the sheriff" as a comparatively unmarked form. In this pair of sample sentences, the marked form "[...] suggests that the agent acted indirectly [...]" whereas the unmarked form indicates that Black Bart caused the sheriff's death in a relatively straightforward and direct way. Grice (1991, p. 37) also provides a nice example: Compare (a) "Miss X sang 'Home Sweet Home'" (an unmarked expression) with (b) "Miss X produced a series of sounds that corresponded closely with the score of 'Home Sweet Home'". If a competent speaker utters (b) in a situation where (a) might have been at least not entirely out of place, the utterance meaning in an instance of producing (b) will be significantly different from producing (a) in the same situation although (a) can be generally seen as mere paraphrase of (b) since both expressions cover the same semantic ground as Horn (cf. 2006, p. 16) put the matter in the quote at the beginning of this footnote.

meaning in contrast to literal sentence meaning. It is the whole tradition [9] of contingent and conventional facts about how we use language and what counts as a normal as opposed to an extraordinary situation, which helps us to make exactly the point we wish to make (i.e., bring utterance meaning into agreement with speaker's meaning) without the need to assemble literal sentence meaning until we finally manage to explicitly exclude all possible but irrelevant (because non-salient) misunderstandings.

What we need in order to be able to exploit this highly efficient way of communicating is just a considerable amount of common background. To say it in Kannetzky's (2001, pp. 204 f)[10] words:

> Such a common background is also needed for the [P]rinciple of [E]xpressibility to work. With regard to the background, the situational meaning of a speech act (that is the utterance meaning, the resulting commitments and entitlements) can be captured by the hearer. Only such a shared background offers the possibility for using "old" linguistic means in a new manner *and* being understood, because it restricts the possible uses and projections of the expressions. Therefore, analogies and metaphors can be exact expressions as well as literally used expressions and are not to be regarded as cases of "inexact language use". The embedding in a certain background also specifies the meaning of vague expressions and secures the usability of the same expression for various purposes, that is it secures the "flexibility of language". [Footnote omitted]

Common background also allows us to convey metaphorical meaning by using language figuratively. Although sentence meaning is always *literal* sentence

[9] Here, the same holds true as for footnote 5 on page 232: If the kind of tradition I refer to is already included in the context of the utterance, just take the following to be redundant and the sentence preceding this note to be false since it *is* only context which enriches utterance meaning.

[10] I think that Kannetzky (2001) clearly draws the same (or at least a rather similar) important distinction between utterance meaning and sentence meaning, as well as between utterance meaning and speaker's meaning. He merely does not give the distinction such a prominent place in his text as I do here.

meaning (as Searle rightly insists; cf. Searle, 1979b, p. 86), utterance meaning may be literal or metaphorical. Since it is sufficient to bring speaker's meaning into agreement with utterance meaning in order to exactly express what we have in mind (and thereby save the Principle of Expressibility), it does not need to be possible that a metaphor can be paraphrased away, even in principle – notwithstanding the fact that the literal meaning of most metaphorically used expressions is quite far from the intended speaker's meaning.

In summary, Searle seems to think that the Principle of Expressibility requires that it must in every case be possible to capture any intended meaning exactly in (literal) sentence meaning. I claim that this is not required by the Principle of Expressibility. It is sufficient if we can bring *utterance meaning* into exact agreement with speaker's meaning. The requirement that we can bring sentence meaning into exact agreement with speaker's meaning is much stricter and – I hold – not necessary for the Principle of Expressibility to come out true.

9.1.4 Searle and Metaphor

We saw that the Principle of Expressibility is not in danger even if we waive Searle's presupposition that every content can be expressed in a non-metaphorical way. By showing this, we saw that the Principle of Expressibility can be upheld separately from Searle's account, i.e., independently of his presupposition that every content can be given a literal form of expression. What then of the internal situation regarding Searle's theory? Is there not an obvious tension between Searle's insistence that every content can be given a literal expression on the one hand and his concession on the other hand

> [...] that we feel that metaphors somehow are intrinsically not paraphrasable. They are not paraphrasable, because without using the metaphorical expression we will not reproduce the semantic content which occurred in the hearer's comprehension of the utterance.
>
> The best we can do in the paraphrase is reproduce the truth conditions of the metaphorical utterance, but the metaphorical

> utterance does more than just convey its truth conditions. It conveys its truth conditions by way of another semantic content, whose truth conditions are not part of the truth conditions of the utterance. The expressive power that we feel is part of good metaphors is largely a matter of two features. The hearer has to figure out what the speaker means – he has to contribute more to the communication than just passive uptake – and he has to do that by going through another and related semantic content from the one which is communicated. (Searle, 1979b, pp. 114 ff)

Searle provides an answer to solve the apparent conundrum in those final sentences of his (1979b) paper on metaphors. A first critical assumption Searle makes is that it is sufficient for an adequate paraphrase to recreate the truth conditions of a metaphorically made assertion. This can certainly be doubted, but it shows that the semantic residue which is not captured by the paraphrase merely plays a role in the hearer, i.e., in metaphor *understanding* and not in its production, according to Searle. Given this setting, it is quite plausible to assume that, for Searle, what we feel is lost in the paraphrase of a metaphor is not part of its content. What gets lost in the literal paraphrase is merely the *effect* a metaphorical utterance can yield in the hearer. This effect is due to the operations the hearer needs to go through in order to understand a metaphorical utterance. This effect is lost in the literal paraphrase of a metaphor. The steps which are needed to understand a metaphor are not necessary to understand its literal paraphrase although it is (truth conditionally speaking) semantically equivalent with the metaphor.

For Searle, the semantic surplus content created by metaphorical utterances, which cannot be achieved via literal paraphrase of metaphorically used expressions, is therefore not part of the intended (speaker's) meaning. It does not concern *what* is being communicated but merely *how* the intended content is expressed. The fact that certain effects cannot be achieved with (literal) language is unproblematic for the Principle of Expressibility. Given Searle's first qualification of the Principle of Expressibility (discussed on pp. 146 f), there is no need for Searle to see the Principle of Expressibility as being endangered by certain effects of metaphorical expressions which

cannot be reproduced in literal language.

9.1.5 The Dilemma for the Challenge to the Principle of Expressibility, Coming From Metaphor

We can, in the end, provide a defense of the Principle of Expressibility against the challenge coming from metaphors by stating the following logical dilemma: Either metaphorical content can always be adequately paraphrased in literal language (Searle's claim) or metaphorical content cannot always be adequately paraphrased in literal language (based, e.g., on Lakoff's conceptual theory of metaphor). If metaphorical content can always be adequately paraphrased in literal language (Q), then the Principle of Expressibility can stay in place, as far as metaphor is concerned (P).[11] If metaphorical content cannot always be adequately paraphrased in literal language ($\neg Q$), then the Principle of Expressibility can also stay in place, as far as metaphor is concerned (P).[12] Therefore, we can conclude that whatever is actually the case regarding metaphorical content – *viz.*, whether it eludes literal paraphrase or not – the Principle of Expressibility can stay in place. For ease of exposition, the dilemma formally looks as follows:[13]

$$Q \vee \neg Q$$
$$Q \rightarrow P$$
$$\neg Q \rightarrow P$$
$$\therefore P$$

It needs to be emphasized again that *"P"* does not stand for the propo-

[11] I think that no one ever doubted this conditional ($Q \rightarrow P$), so no justification for it is needed.

[12] This is because assuming that every content needs to be expressible in literal sentence meaning for the Principle of Expressibility to come out true is a mistake. That we can exactly capture any intended (metaphorical) meaning in utterance meaning is sufficient for the Principle of Expressibility to hold true, as far as metaphor is concerned. Therefore, the conditional ($\neg Q \rightarrow P$) is also true.

[13] Where the first premise ($Q \vee \neg Q$), being a logical truth, is not even necessary to yield a valid argument form.

sition that the Principle of Expressibility can stay in place, full stop. A conclusion that strong can only be derived when all apparent problem cases for the Principle of Expressibility are ruled out. We are not there yet, obviously. What *"P"* stands for here is, as indicated above, that the Principle of Expressibility can stay in place, *as far as metaphors are concerned.* In other words, we only know that the truth or falsity of Q – i.e., the question whether metaphorical content can always be adequately paraphrased in literal language – is not decisive for the truth of the Principle of Expressibility. Therefore, metaphors are ruled out as problem cases for the Principle of Expressibility. We will soon find ourselves confronted with an analogous dilemma regarding other kinds of allegedly inexpressible content on pp. 250 ff.

9.2 What Mary Could Not Say

9.2.1 The Problem

Another aspect which might seem to be in conflict with the Principle of Expressibility is the apparent ineffability of experiences. Often not even poetry seems equipped to give adequate expression to the more emotional side of the human condition: love, fear, kindness, hate, weltschmerz, and so on. How could language ever communicate the profoundness of such feelings? Analogously, we often feel a deep inadequacy of language to properly express what impact certain experiences may have on us. How could language alone ever make it possible to capture the impression of an especially sublime sunrise? Or how could we ever explain what it is like to walk barefooted towards a rainbow in summer rain? Does not expressing experiences like these obviously transcend the capabilities of language by far?

9.2.2 What Mary Has to Do With Language

I think that any tendency to reject the Principle of Expressibility on the basis of such considerations is built on a severe misunderstanding. Let us consider, in comparison to the previously mentioned examples, a somewhat more pro-

fane but well-established philosophical case in point: Frank Jackson's (1982) famous Mary.[14] The argument involving Mary, the specialist about vision who has never seen a color in her entire life, is widely known. So we can be content with just a quick recapitulation of her story:

> Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational fats consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. (Jackson, 1986, p. 291)

The thought experiment about Mary is concerned with refuting physicalism by means of the knowledge argument (cf. Jackson, 1982, pp. 128 ff): If physicalism was true, Mary would not learn anything new by seeing color for the first time. This is what the case of Mary is supposed to show,[15] but how can Mary serve our discussion of the Principle of Expressibility?

Well, if we accept Jackson's claim that Mary learns something new when she first leaves her monochrome environment, we can also conclude that, if exclusively linguistic expression was capable of explaining to someone what it is like to see, e.g., a ripe tomato, then the knowledge argument would fail – or at least not get off the ground. Mary could have learned what it is like to

[14] I will explain how Mary's case can serve our discussion in a moment. A crucial aspect to remember is that we are not merely concerned with the Principle of Expressibility as Searle endorses it, according to which everything that can be thought can also be said, but with the *strengthened* Principle of Expressibility, according to which whatever can be thought can be expressed *and* understood by others. (See the statement of the strengthened Principle of Expressibility on page 150, and the surrounding discussion in sections 6.2.4 and 6.2.5.) If not indicated otherwise, it is always the strengthened version of the Principle of Expressibility which I discuss here although I usually waive adding the explicit qualification "strengthened" each and every time I mention the principle.

[15] For a quick overview and explanation of the argument and the argumentative background, see, e.g., Torin (2010) and Horowitz (2011).

see color from her black-and-white screen by reading pertinent explanations if language was in principle able to convey this information.[16] She has all the information which can be linguistically conveyed but still learns something new when she is first confronted with a colored object, according to the knowledge argument. This seems to strongly suggest that there is content – namely qualitative content – which cannot be communicated. Does this prove that the Principle of Expressibility is wrong?

Given the characterization of thought in section 1.3 on page 11 as everything that can pass through one's stream of consciousness,[17] the qualitative aspect of experiences undoubtedly belongs to thought. Although I have a slight preference for a representational account of qualitative character, as advocated, e.g., by Tye (1996) or Harman (1990), we should presuppose a non-representational account of phenomenal character at this point, as discussed in Block (2007), since the latter appears to be more challenging for the Principle of Expressibility.[18] As the Principle of Expressibility claims that everything that can be thought can also be communicated, the assumption that qualitative aspects of experience cannot be communicated seems to lead to a straightforward refutation of the Principle of Expressibility.

However, we need to remember Searle's first qualification of the Principle of Expressibility, discussed on pp. 146 f, where Searle reminds us that

> [...] it should be emphasized that the principle of expressibility does not imply that it is always possible to find or invent a form

[16] We might need to slightly adapt Jackson's description of the case, so that Mary learns everything which can be linguistically conveyed (about color). For the physicalist, this will not make much of a difference since everything that can be linguistically conveyed (about color) will be contained in what Mary learns, namely "[...] everything there is to know about the physical nature of the world." (Jackson, 1986, p. 291) For the dualist, however, the aforementioned adaptation might be crucial to exclude that Mary could have learned about qualitative character before she left her room, but her curriculum was simply incomplete because it exclusively focused on the *physical* nature of the world.

[17] See especially footnote 11 on page 11.

[18] At least if we assume that qualia, as the non-representational phenomenal aspect of experience, still count as mental *content*. If phenomenal character does not count as content, it drops out of the scope of the Principle of Expressibility, which claims that every *content* that can be meant or thought can also be expressed. If, of course, qualia do not qualify as content, they also cannot belong to thought. But then, I presume, we also could not be aware of qualitative aspects of experience.

> of expression that will produce all the effects in hearers that one means to produce; for example, literary or poetic effects, emotions, beliefs, and so on. We need to distinguish what a speaker means from certain kinds of effects he intends to produce in his hearers. (Searle, 2011, p. 20)

Trying to evoke qualitative effects in a hearer is a paradigm example of this first and principal qualification Searle adds to his introduction of the Principle of Expressibility. Trying, so to say, to "transplant" an experience with all its qualitative aspects into the mind of a different person by means of linguistic communication is definitely trying to achieve a certain *effect* in the hearer. So, trying to explain to Mary what it feels like to see a ripe tomato by making her have a pertinent experience is an attempt to produce a certain effect – namely Mary's having an experience of what it is like to see a red object – and must therefore count as a pragmatic aspect of communication.[19] Searle is very clear about the fact that these effects are not covered by the Principle of Expressibility. A failure to produce certain effects via linguistic expression – even if it is a principled failure which cannot be overcome – is therefore no reason to reject the Principle of Expressibility.

[19] Apart from these considerations, there is an additional problem with the idea of "transplanting" a qualitative experience into somebody else's mind: If we wish to put the experience of what it is like to, say, see something red into Mary's mind, then it is unclear from (or with?) which "perspective" the experience should be transplanted. Should Mary receive the experience of what it is like to see something red *for the "donor"* of the experience, or should she receive the experience of what it is like to see something red *for a human being*, or should she receive the experience of what it is like to see something red *for her*? It might be possible to find a convincing answer to this question, but the correct (or at least the most plausible) answer is not obvious, or so I claim. I suspect that the entire idea of "transplanting" an experience is confused and wrongheaded from the start. Nevertheless, it can serve well to point out and explain certain confusions which may underlie a temptation to reject the Principle of Expressibility for the reasons presently considered.

9.2.3 What the Principle of Expressibility Requires Regarding Mary, and What Not – Part 1: Irreducibly Phenomenal Content Denied

We need to be careful not to conclude that there are indeed things which cannot be communicated because we cannot "persuade" Mary to know what it feels like to see a ripe tomato. This is not a case where we fail to adequately express what it feels like to see a red object. We can perfectly well talk about all aspects of such an experience, including its qualitative aspects. It is also not a case where we can say something which cannot be understood, for Mary may perfectly understand all our descriptions of what it feels like to see color. There is no misunderstanding involved since Mary knows all the expressions we use and can follow our description of the experience. It would therefore be incorrect to say that she does not *understand* what we say. What makes our attempts to tell her what it is like to see color appear deficient is our inability to evoke the corresponding experiences in Mary. We cannot make her know what it feels like to see color (from the first-person perspective) because she simply lacks the relevant experience.

In order to know (in one sense of "to know") what it is like to have a certain experience, one needs to live through the experience on one's own. This is not only the case for sensual experiences, like seeing color, but might also be the case for purely intellectual contents. If there is in fact something like cognitive phenomenology, then thinking that P (in the sense of believing that P is true) will "feel" differently from thinking that Q, and both will "feel" differently from assuming, conjecturing, or doubting that P, or that Q (cf. Chudnoff, 2015; Voltolini, 2017; and several contributions in Bayne & Montague, 2011a). Even if there is no doubt that P and Q can be adequately expressed, it will remain dubitable that one can explain what it is like to assume or doubt that P.

There is, of course, a noteworthy ambiguity present in the expressions "explain" and "know" regarding the present context. An able phenomenologist will eventually find the right words to explain what it is like to doubt some-

thing, in contrast to conjecturing or assuming something.[20] The phenomenologist's description, given that there is no linguistic misunderstanding, can be understood by someone who has never doubted or assumed anything. In this case, the person in question does know (again, in one sense of "to know") what it is like to doubt something. He has followed the honest and accurate description of the phenomenologist and therefore has an appropriate depiction at hand of what it feels like to doubt something. In this sense, our curious never-doubter now knows what it is like to doubt something. Still, in another sense of the term "to know" it remains doubtful whether the person in question really knows what it is like to doubt something.[21]

We may overcome the cumbersome wordings present in the previous paragraph by borrowing Bertrand Russell's notions of 'knowledge by acquaintance' and 'knowledge by description' (cf. Russell, 1911).[22] It might be conceded that it is possible to acquire knowledge by description of qualitative aspects – be it of sensual experiences or of cognitive phenomena – by being

[20] What qualifies our able phenomenologist is a sublime talent to provide enlightening descriptions of experiences in a non-question-begging way for subjects who never had the experiences in question themselves. Our able phenomenologist probably does not achieve that by introducing apt technical terms but via skillful usage of everyday language expressions. So, what qualifies our able phenomenologist might be a literary aptitude rather than philosophical competence or – I prefer to say – probably a combination of both. The point is that we do not necessarily need anything that goes beyond everyday language competence in order to be an able phenomenologist, but it should nevertheless be acknowledged that coming up with apt phenomenological descriptions is not a trivial task to be completed.

[21] Presumably because the never-doubter – pace Michael Tye (2009) – lacks certain phenomenal concepts which he cannot acquire unless he stops being a never-doubter, i.e., before he actually doubts something for the first time.

[22] Note that Russell might not have been entirely satisfied with this application of his distinction because, first of all, I assume that we can be acquainted with material or physical objects, not only with sense data – pace Russell (1911, pp. 112 and 127). Secondly, in order to be acquainted with phenomenal aspects of experiences, it seems necessary to include experiences (which have or carry phenomenal aspects) in our ontology of mental entities as something that Russell (1911, pp. 119 and 128) might have called an "idea," namely a mental representation of something that is (or at least might be) not mind-dependent. I wholeheartedly agree with Russell (1911, p. 119) that ideas must not "[...] become a veil between us and outside things [...]", but it is nevertheless unclear whether Russell would be happy with the claim that we can also be acquainted with (phenomenal aspects of) ideas in this sense – notably not only experiences of seeing, hearing, tasting something, etc. but also (potentially) cognitive experiences like in the case of our never-doubter on the current page and below on the facing page.

presented with a good linguistic depiction of the aspects in question. But knowledge by acquaintance, so an objection might go, cannot be conveyed by means of language alone in many cases. Since thought in the explicitly adopted wide meaning of the term (see footnote 11 on page 11) comprises also qualitative experiences and thereby qualitative content which cannot be adequately communicated, the Principle of Expressibility fails because there are mental contents which cannot be adequately expressed.

One way to answer this objection amounts to a kind of defense prominently deployed by some physicalists against the knowledge argument (cf., e.g., Lewis, 1999; and Conee, 1994): We might say that knowledge by description and knowledge by acquaintance of a certain thing are merely two different ways of acquiring information about the same thing. If I read about, say, Barack Obama and if I meet him in person, I acquire information about the very same person. I acquire, let us assume, the very same information – for example how tall Barack Obama is – merely in two different ways. Yet, if I have already gathered the information about how tall this person is from a textual description and then see how tall the person is in real life, I do not acquire new information. The physicalist's defense would have it that I learn about the very same fact I knew already, albeit now in a different way.

Coming back to our discussion about experiences and their qualitative aspects, we can analogously say that Mary did not learn anything new when she first saw color. She already had the relevant information from her studies. What she learns is the same information – namely what it is like to see a ripe tomato – given in a different way. The same holds true regarding our never-doubter on the facing page: The person already knows what it is like to doubt something from the phenomenological description he was provided with. When the person in question now actually doubts something for the first time, he will merely achieve a new way of knowing what it is like to doubt something. He will not acquire new information since he had the information already. The information he already had is merely presented in a new way: a way in which he did not have access to this information before. He now has first-person perspective knowledge of what it is like to doubt something while he only had third-person perspective knowledge of the very same thing

before.

In addition, the never-doubter and Mary acquire new abilities, namely the ability to remember, imagine, and recognize the experience of doubting something and of seeing something red respectively. But learning a new ability (e.g., to remember, imagine, or recognize certain experiences or certain contents; cf. Lewis, 1999, pp. 286-290) is not the same thing as acquiring new content. The content might already have been there; just certain abilities regarding this content (e.g., to remember, imagine, or recognize it) are newly acquired. So knowledge by description – the materialist might say – provides, in principle, access to the very same content we can also acquire via being acquainted with a certain phenomenon. Seeing the sun and reading or hearing about the sun concern the very same object and can provide the very same information – just not in the same way, and in consequence not in connection with the same abilities.

We might also say that knowledge by description (i.e., linguistically conveyed knowledge) provides us with "theoretical" knowledge – a case of *knowing that* – whereas knowledge by acquaintance provides us with first-person "practical" knowledge – a case of *know-how*. The Principle of Expressibility is not threatened by making this distinction because everything we can have know-how of we can also acquire know-that about by means of linguistic communication even if we cannot acquire know-how (i.e., abilities; cf. Lewis, 1999, p. 288) merely by being provided with know-that (i.e., propositional information) in most cases. Even supposing that the *Ability Hypothesis* (cf. Lewis, 1999, pp. 285-290) or the *Acquaintance Hypothesis* (cf. Conee, 1994) of different kinds of access to the very same information are not accepted, the Principle of Expressibility does not need to be abandoned. We can still take recourse to the distinction emphasized by Searle: namely the distinction between what is said and what effect can be achieved with what is said.

9.2.4 What the Principle of Expressibility Requires Regarding Mary, and What Not – Part 2: Irreducibly Phenomenal Content *Sui Generis* Acknowledged

It is quite plausible to assume that we can make the qualitative aspects of experiences the topic of our communication even if it might be doubted that we can – even in principle – find adequate means to express these aspects. That we can talk about the qualitative aspects, even if we cannot adequately express them, is something I take as a given. Some of us address this part of our lives in language more frequently and more willingly than others, but we all do it at least occasionally, I would say. We share our impressions of what it was like to experience this or that, how it felt to live through this and that situation, and so on. So, qualitative aspects of experience can be linguistically addressed. Let us assume, by way of rejecting the Ability Hypothesis and the Acquaintance Hypothesis mentioned before, that knowledge by acquaintance gives us access to different content than knowledge by description, not only different access to the same content. Does the Principle of Expressibility need to be rejected on the basis of this assumption?

I do not think that this is the case because we can still argue that the deeply felt inadequacy of linguistic expression concerning qualitative aspects is not due to the fact that there are things we cannot adequately express. Rather, this impression comes from an occurrent inability to always achieve the desired effects with language in addressees even if the linguistic expression is perfectly adequate. What might this defense of the Principle of Expressibility look like? In a nutshell, it looks quite similar as in our discussion regarding metaphor in section 9.1.

We can say that getting someone to know what it is like to see red or doubt something from the first-person perspective is an *effect* we wish to achieve with linguistic explanations. That this effect cannot be achieved if the addressee of our explanations has never experienced the addressed phenomenon is, taken in this sense, no problem for the Principle of Expressibility. Just as our inability to convince someone of a bad idea (as discussed on page 147) does not indicate a lacuna in our language which is in want of being filled, so

is our inability to "transplant" a mental content from one consciousness into another not a shortcoming of the expressive powers of language. For most experiences it holds that we cannot simply talk people into having them. If you have never lived through a certain experience, then you lack acquaintance with what it is like to have a certain experience. No string of text could ever change that. For all that, these "deficiencies" concern merely the ability to achieve certain effects in a hearer, not a principled inability to adequately frame a given content in language. In other words, the failure is merely pragmatic or perlocutionary, not semantic. Since the Principle of Expressibility concerns only what can be said but not which effects can be achieved, as Searle emphatically makes clear (see pp. 146 f), our inability to make someone have an intended or meant experience does not contradict the Principle of Expressibility.

Regarding Jackson's thought experiment this means that we have several ways to bring the Principle of Expressibility into agreement with what happens to Mary. Either we can side with the physicalist and claim that Mary merely learns about old information in a new way. Or we can side with the dualist and concede that Mary acquires substantially new information, i.e., information she did not have and could not have acquired in her monochrome environment before she went through the relevant experience herself. The Principle of Expressibility can, I maintain, successfully be defended in both cases. So, we have the same kind of dilemma regarding the challenge to the Principle of Expressibility coming from qualitative content, which we have already encountered regarding metaphors in section 9.1.5:

In the former case, assuming that physicalism is true, we can say that language merely presents in a different guise the same qualitative aspects, which we can also learn about by undergoing certain experiences. Since it is the same content we can experience and converse about, there is no content left which cannot be expressed. In the latter case, assuming that dualism is true, we can explain away the perceived inadequacy of language to communicate qualitative aspects by pointing out a crucial distinction which Searle made right away when he presented his version of the Principle of Expressibility: What we actually feel language is insufficient for is not expressing qualitative

content adequately – i.e., in a way which allows other people to grasp exactly what I wish to express, given that they have the required (phenomenal) concepts – but our inability to achieve certain effects in hearers. Achieving our aims in such a case could only be realized with the help of telepathy or magic, if at all. It is clear, I hope, that the Principle of Expressibility should not be understood in a way which requires telepathy or magic to be real, or even possible, for the principle to be true.

What I called "transplanting" a given content into another mind directly – i.e., without a communicative medium – is what language cannot achieve, even in principle. Linguistic communication is neither telepathy nor magic, so we cannot make someone have an experience which this person has never had. All the same, this is not a shortcoming of linguistic communication. The problem, we might say, ought rather to be sought in the expectation that language would enable us to achieve something which perhaps not even telepathy or magic would allow us to do. Chastening our expectations in this regard can already have a mitigating effect on the perceived challenge Mary's and similar cases might pose for the Principle of Expressibility.

We can – in principle – give adequate expression to every possible content we can think or mean, and we can do so – in principle – in a way which allows other people to understand what we say. That not *everyone* might be in a position to understand what we give expression to – even if the content was adequately expressed – is not in conflict with the Principle of Expressibility, not even with its strengthened version. That we cannot "transplant" mental content into another mind is not due to the fact that there is something we can think – *viz.*, that there is mental content we can have – but not adequately express. That we can in fact give adequate expression even to qualitative content is plausible if we assume that a hearer who has lived through pertinent situations and has therefore had the relevant experiences is indeed able to perfectly recreate, on the basis of our description, the content we wish to express.[23]

[23] Or such a hearer might at least be able to realize that he does not have memories of the pertinent experiences available which would be needed to recreate the mental content we wish to give expression to. A good description of (what it is like to live through) a certain experience certainly allows the hearer to understand whether he has ever had said

That we cannot explain to a blind person what it is like to see something red is not an imperfection of language, not a lack of expressive capacities, but an impossibility to achieve certain *effects* with language. Having this kind of expectations of language's expressive power is, I think, generally wrongheaded and unwarranted. Yet it is still not an expectation that would – even if we are unwilling to abandon these unrealistic expectations – license rejection of the Principle of Expressibility. At least it will not if we respect the crucial difference between what can be said (which is relevant for any attempt to refute the Principle of Expressibility) and which effects can and cannot be achieved via linguistic expression (which is utterly irrelevant for any challenge posed to the Principle of Expressibility).

It is not the case that we try to say something which cannot – even in principle – be adequately expressed and understood by another person when we attempt to give expression to qualitative or phenomenal aspects of our experiences. If somebody could recall the relevant experience, she would perfectly understand what we were talking about – namely how it feels or what it is like to see color, for example. The subjective character of what it is like to undergo a certain experience is therefore not beyond the reach of linguistic explanation. If your interlocutor has the necessary background to empathize with your depiction and is willing to recreate your impression of a given experience in her imagination, then I see no good reason why we should assume that it is in principle impossible not only to tell someone what going through a certain experience was like, but also to be perfectly understood by the other person.[24]

experience. If misunderstandings regarding the correct identification of (what it is like to have) a certain experience can be avoided, this should count as a good indicator that the qualitative aspect in question was indeed adequately expressed.

[24] Of course, it might be impossible to ever *know* whether the other person really correctly understood which phenomenal content you wished to communicate. But this is an epistemic problem, and concluding that we cannot communicate certain content because we can never know whether we really achieved our aim is just the sort of illegitimate conflation of epistemic with metaphysical considerations I called out in section 8.3.2.3. (Except, of course, if we could provide a cogent argument to the effect that just because we can never *know* whether we are properly understood, we are licensed to conclude that we can never *be* adequately understood in relevant cases, i.e., when we are concerned with qualitative aspects of experience. At any rate, I honestly cannot even imagine what a sound argument to this effect could possibly look like.)

Language is generally able to reach much further than we often assume, if we overcome idiosyncratic and contingent limitations of individual eloquence. Even domains which traditionally often count as ineffable are not beyond the reach of being expressed in language.[25] If you can think it, then you can – at least in principle – also say it, and others can – at least in principle – understand it. The Principle of Expressibility is not threatened from this side. And regarding Mary, there are things she does not know simply because there are certain experiences she never made. Still, there is nothing we might want to tell Mary that she cannot in principle understand. Even more importantly, there is nothing we or Mary can think or mean but not say, at least in principle.

[25] Among the most historically prominent candidates to be mentioned here are certainly ineffability claims coming from religious experience and mysticism. Notwithstanding that also "[...] philosophy of mathematics, and contemporary cognitive science" (Kukla, 2005, p. 1) are a source of ineffability claims. Addressing these in detail is beyond the scope of the present investigation, but note that not all ineffability claims are in conflict with the Principle of Expressibility. Claims to the effect "[...] that there are bound to be some hypotheses that the human mind is incapable of entertaining" (Kukla, 2005, p. 2) do not threaten the Principle of Expressibility. The principle only claims that whatever can be thought can also be said, but the Principle of Expressibility does not concern what cannot be thought. Only "[...] the claim that we are able to understand or come to know certain truths which it is beyond the power of language to express" (Kukla, 2005, p. 1) cannot be upheld if the Principle of Expressibility is true. This distinction must be crucial for any further investigation of ineffability claims and the Principle of Expressibility, but we have to leave the matter at this preliminary remark.

Chapter 10

The Legitimacy of Unexpressed Meaning

The last sentence of the previous section would have been a worthy final claim for this last part of the investigation at hand. Unfortunately, most of what we achieved so far stands on shaky ground as long as we have not settled that unexpressed and therefore merely mental content is a theoretically warranted category. This critique of the strategy, which the entire Part III builds on, has loomed in the background since we first encountered the claim that unexpressed meanings are problematic in section 6.2.6.1 on pp. 154 f.[1] The objection against presupposing the existence of merely mental (and potentially inexpressible) content is, to repeat, on the basis of Quine's dictum (already mentioned on pp. 154 and 191), that there is no guarantee that we can provide identity criteria for this kind of content. As long as identity conditions for a kind of entity are dubious, so the challenge proceeds, the kind of entity in question must not be adopted in the ontology of a theory. This critique threatens to undermine the integrity at least of set M which contains whatever can be meant or thought – regardless of whether it can

[1] I claimed on page 160 that Binkley's critique is no threat against the methodological use Searle makes of the Principle of Expressibility, for explicitness (which Searle relies on) is a kind of precision, not accuracy, and therefore does not presuppose a comparison of expressed with unexpressed meaning. Nevertheless, my application of the principle is potentially vulnerable to the point raised in Binkley (1979) since I do claim that whatever can be meant can be accurately (and comprehensibly) expressed.

also be expressed. Even set S is vulnerable to this rationale since S comprises what *can* be said and therefore also potentially includes unexpressed content. The very set theoretic basis of the investigation at hand is therefore endangered if we cannot secure the theoretical legitimacy of the notion of unexpressed meaning. Filling this gap is the purpose of the current chapter.

10.1 No Thought Without Talk?

We can also frame the concern regarding unexpressed (and potentially inexpressible) mental content in a less theoretically rigorous way: Building on Heinrich von Kleist's (2004b) 'On the Gradual Production of Thoughts Whilst Speaking',[2] it takes but a slight exaggeration of the position suggested in this short essay to come to the conclusion that thoughts can be properly developed, if not exclusively while speaking or writing, at least only in language. On the basis of this position we should conclude that every thought – at least if it is sufficiently developed to even count as a thought – is linguistically framed. Since every linguistically framed thought must trivially be expressible in language (at least in principle), there simply is no merely mental and potentially inexpressible content. Everything which deserves to be called "content" at all is therefore linguistic content, and a set of merely thinkable content (M) which might contain meanings which cannot be communicated is either illusory and ill-framed or superfluous because there simply is no content which could be sorted into M but not also into S.

This means that content does not need to be expressed already, but in order to ascribe a certain mental content, at least to a linguistic creature, the creature in question needs to know how to express the intended content. Otherwise – i.e., in the case of merely mental and at least potentially inexpressible content – we cannot ascribe any mental content due to lack of

[2] The original German title is 'Über die allmähliche Verfertigung der Gedanken beim Reden' (cf. von Kleist, 1924). The text was probably written in 1805-6, but published only posthumously in 1878 (cf. von Kleist, 2004a, p. 440). The German text is easily available via `http://hdl.handle.net/11858/00-001M-0000-002B-B33A-4` or `http://www.kleist.org/index.php/downloads-u-a-werke-im-volltext/category/16-heinrich-von-kleist-aufsaetze`.

identity criteria for the content to be ascribed. So, if the content cannot even be identified and individuated in principle because it is indeed inexpressible content, we can conclude that there is no content at all, for there is no content which can be ascribed.[3]

The position just sketched is admittedly rather crude and needs to be developed and refined in order to have any chance of being somewhat convincing. Despite its crudeness, I confess that I have a good deal of sympathy for the basic tenet which underlies such a position. Drawing on the fusion of Wittgenstein's dictum – that whatever can be said can be said clearly – with the (strengthened) Principle of Expressibility – according to which there is no content which principally cannot be adequately and exactly expressed (and understood by others) – we can conclude that there is no content which cannot be expressed *clearly*. In consequence, where only a confused expression is possible, there was no content to begin with. Yet this would be a clear over-

[3] It needs to be emphasized again that this is not the position von Kleist endorses in his text. He, to the contrary, claims that

> [...] if an idea is expressed confusedly we should by no means assume that it was thought confusedly too; on the contrary, it might well be the case that the most confusedly expressed ideas are the clearest thought. (von Kleist, 2004b, p. 408)

I tend to think, pace von Kleist, that confusedly expressed ideas are a good indicator of confused thought – except where insufficient language proficiency, impaired speech, or similar conditions suggest that only communication but not the underlying thoughts are affected by confusion. However, in cases where no such exculpation is available and where a speaker is unable to clearly spell out his line of thought in any communicative system, a confused expression of ideas is usually a quite reliable indicator of confused thinking (or reasoning).

Kleist (2004b), in any case, thinks of language (or maybe only of speech) not as a medium of but, rather, as a catalyst for thought. This is quite evident from the very beginning of his essay, which he starts with the following lines:

> If there is something you wish to know and by meditation you cannot find it, my advice to you, my [dear reader], is: speak about it with the first acquaintance you encounter. He does not need to be especially perspicacious, nor do I mean that you should ask his opinion, not at all. On the contrary, you should yourself tell him at once what it is you wish to know. (von Kleist, 2004b, p. 405)

After having provided several examples for the stimulating effect of speaking (or generally of using language?) on thought, he concludes that: "Speech then is not at all an impediment; it is not, as one might say, a brake on the mind but rather a second wheel running along parallel on the same axle." (von Kleist, 2004b, p. 408)

reaction, which I blame on the mistaken (and apparently again somewhat verificationist) assumption that if we cannot precisely identify the content to be ascribed, then there was no content in the first place. We should at least consider the possibility that a confused or vague expression of ideas might be an adequate formulation of vague or confused thinking.

10.2 Evidence for Unexpressed Content

However, regarding merely mental (and potentially inexpressible) content,[4] I think that there are certain everyday phenomena which should convince us to let go of theoretical scruples à la Quine and accept that unexpressed meanings should also find their place in a solid theoretical investigation of the relation between mind and language.

The tip of the tongue phenomenon: One of these phenomena is the tip of the tongue phenomenon, quickly mentioned already on pp. 36 f. When we have something on the tip of our tongue, we feel that we exactly know what we wish to say but cannot retrieve the appropriate expression for the content to be expressed. This phenomenon is extremely common, and its existence is universally accepted (cf. Brown, 2012, pp. 5 and 28). Although the tip of the tongue phenomenon provides good evidence for the existence of unexpressed mental content, albeit of course not for inexpressible content, I wish to quickly address a probably kindred but still different phenomenon in addition.

[4] I will not provide any evidence for actually inexpressible content – in contrast to expressible but actually unexpressed content – since I do not think that there is good evidence for inexpressible content and, consequentially, I also do not think that there is inexpressible content. The reason why I nevertheless repeatedly mention at least potentially inexpressible content is that I take it to be important to have at least the conceptual means to discuss the question whether there might be truly inexpressible content even if in fact there is no such content. Cutting oneself off from the conceptual resources to even address the question/possibility of whether/that there might be inexpressible content is, as demonstrated in section 7.2 (and especially in section 7.2.2), certainly to be avoided.

The student-professor case: The situation I have in mind is one where we struggle to put our thoughts into words and indeed achieve a somewhat acceptable result of expressing what we have in mind, but we later come across somebody else's formulation and think: "That's exactly what I meant!" I assume that this kind of situation is, although probably not as widespread as the tip of the tongue state, still fairly well-known. In order to better illustrate what I have in mind nonetheless, here is some anecdotal evidence for the claim that the phenomenon I have in mind really exists: A former colleague once told me about a professor of hers who was especially gifted in the following respect. For example after a presentation in a PhD colloquium, he tended to summarize the most crucial points of the PhD candidate's talk and usually introduced his summary with the words "So you mean to say that ...". As far as I recall the anecdote, the common reaction of most PhD candidates was to reply "Yes, that's exactly what I meant to say," followed by a silent "And I wish I could have put it that way beforehand!"

What frequently happened – or at least might have happened, I claim – in these situations is that the professor's formulation gave better expression to the content the student aimed to communicate. Importantly, I do not only mean that the professor's way of putting the matter was more pointed, eloquent, compelling, etc., but that the professor's words gave a more *accurate*[5] expression to the content which the student wished to communicate than the student's own formulation. In order to make such a judgment, the student needs to (be able to) compare his intended but hitherto not (adequately) expressed meaning with the meaning of what the professor said. If we think that such a comparison is even possible, we need to assume that there indeed are unexpressed meanings which should also find a place in our ontology for the present investigation. The phenomena I cited do not make the ontological status of unexpressed meanings less dubious, but they provide good reason for not abandoning them nevertheless.[6] In other words, if there are indeed

[5] In the sense discussed in section 6.2.6.

[6] I adopt a similar attitude as the one expressed in Kripke (cf. 1981, p. 39, n. 11): It is not necessary (although it might be desirable) to have a notion rigorously defined before it can be legitimately used in philosophical (or any other) investigation. I also agree with Kripke (1981, p. 42) that intuitive evidence for a phenomenon can outweigh theoretical

unexpressed meanings, but our methodological principles prohibit that they find their way into the ontology we build our theories on, so much the worse for our theories and our methodological principles.

10.2.1 How to Deal With Quine's Dictum

What I mean to say is that the way Quine's dictum is formulated and usually understood has potential to hamper scientific advance, rather than to improve the quality of investigation by proposing rigorous theoretical scrutiny. We should certainly aim and strive for precise identity criteria. There is no doubt whatsoever about that. But it would be illusory to assume that a scientific (which includes philosophical) investigation can always already start with a perfect understanding of the phenomenon to be studied. A discovery of precise identity criteria for an opaque phenomenon is a high achievement which rather marks the culmination of a successful investigation, but it cannot be presupposed before any inquiry is even allowed to begin. In some cases, the best we can hope for might be a thorough explication of plausible candidates for identity criteria which might serve for a stipulation in order to (primarily) demarcate the phenomenon in question, rather than a true discovery. In any case, identity criteria cannot be required for every investigation beforehand, in order to license an inquiry in the first place.

To be sure, I do not claim that Quine's dictum[7] cannot be interpreted in a way which makes it compatible with the points I have raised. For example, the dictum might be – taken in a (robustly) realistic way – interpreted as claiming that for every (real) entity there *are* clear identity criteria which merely need to be found. This understanding of Quine's dictum could serve to inspire and motivate more thorough investigation. Yet, the way it is

scruples.

[7] Quine's œuvre provides plenty of candidates to go under the label "Quine's dictum." The first to come to mind for most people might not be what is under consideration here, but rather Quine's famous dictum of ontological commitment: "To be is to be the value of a [bound] variable." (Cf. Quine, 1948, pp. 32 and 34; and Quine, 1951a, p. 11) Other candidates for the title "Quine's dictum" might plausibly be available as well. However, when I use the expression "Quine's dictum" here, I always mean his maxim that there is *no entity without identity.*

usually understood, I am afraid, rather contributes to preventing serious investigation in many cases since the phenomena to be studied get prematurely dismissed as dubious or unscientific.

In order to avoid any misunderstanding, I want to emphasize that I do not, of course, plead for including witches, dragons, and God(s) into the ontology of a scientific theory.[8] Nonetheless I do think that nothing – in principle – lies outside the scope of scientific investigation.[9] A principle which prevents scientists from investigating something merely because it is not understood well enough to meet scientific criteria of description yet is certainly not a good methodological principle – given that a good principle would encourage and advance investigation, rather than obstruct it.

10.2.2 Tip of the Tongue and Feeling of Knowing

Let us bring the discussion away from these quite general considerations and back to the topic of unexpressed meanings. The phenomenon depicted via the situation of the professor and the student and the tip of the tongue phenomenon are similar insofar as they both come accompanied by the clear impression that one would recognize the correct words if they were presented. A person in a tip of the tongue state can identify the expression she is momentarily unable to retrieve when the wanted word is presented to her. Similarly, the student can identify the professor's formulation as giving (better) expression to what the student had in mind. One difference between tip of the tongue states and the other situation described is that the former are usually accompanied by a strong feeling of knowing (the word which is temporarily unavailable) without being able to retrieve this knowledge at the

[8] All the same, we clearly need to make room for fictional/mythological dragons, the (folk) concept of witches, and literary traditions concerned with God(s) as topics of scientific investigation. See section 5.1.3 for some remarks about the ontology of fictional entities.

[9] This claim must not be misunderstood as an endorsement of scientism (cf. Stenmark, 2013). On the contrary, I plead for *scientific expansionism* (cf. Stenmark, 2013, p. 2104) with the "twist" that the term "scientific" is here explicitly understood as not only covering the natural sciences but also methodologies from the humanities, as already indicated on the preceding page. I, at least, cannot find any conflict in an attempt to combine methodical rigor with disciplinary and methodological pluralism.

moment.[10] This is typically not the case in the professor-student situation. If the student had known how to properly express what he had in mind, he would have done so. So, there is no feeling of knowing regarding the expression, but there is one regarding what the expression means.

A slight variation of the situation considered might bridge this difference. We can imagine that a person has forgotten a particularly apt formulation of her thoughts. She tries to remember or recreate the expression but fails. She then comes across the words of somebody else which exactly match the formulation she has been trying to remember. In this case, we have a feeling of knowing the words,[11] which makes this case quite similar to a common tip of the tongue situation, but we are not dealing with (hitherto) unexpressed meaning anymore. The meaning has already been expressed even though the expression (but not the fact that there was an apt expression) has been subsequently forgotten. Therefore, this variation is crucially different from the student-professor situation we started out with since in this situation the content in question was not (properly) expressed by the student beforehand and might never have been aptly expressed – but perhaps thought several times already – before the professor provided his formulation.

[10] There is an ongoing debate about whether tip of the tongue phenomena are merely a quite intense version of the more general phenomenon of feeling of knowing (FOK) or whether feeling of knowing and tip of the tongue (TOT) are different phenomena. Brown (2012, p. 20) summarizes the debate, based primarily on neuro imagining techniques, by stating that "[...] it is most likely that TOT and FOK responses are highly related but distinctive cognitive functions." An important terminological difference between TOT and FOK for our purposes concerns the fact that "[...] an essential part of a TOT is a sense that the word eventually can be *recalled* given sufficient time, whereas FOK judgments involve the likelihood of subsequent correct *recognition*." (Brown, 2012, p. 17) "An FOK judgment is a general assessment of one's sense of familiarity for inaccessible information [...]" and "TOTs and FOKs are similar in that both relate to unavailable knowledge." (Brown, 2012, p. 16) However, the intuitive understanding of tip of the tongue phenomena includes a high degree of feeling of knowing. Test subjects will generally follow this natural understanding of TOT as including FOK in laboratory settings unless they are explicitly instructed to keep them separate (cf. Brown, 2012, p. 8). I will also follow this intuitive understanding of what a tip of the tongue state amounts to and therefore treat TOTs as involving FOKs.

[11] But no FOK regarding the content since the content is always readily available throughout this imagined scenario. As explained in footnote 10, FOKs only relate to inaccessible information or unavailable knowledge.

10.2.3 Explaining Away Unexpressed Meanings?

I cannot rule out the possibility that phenomena of the kind I have cited in favor of the actual existence of unexpressed meanings can be explained away as mere déjà-vu experiences.[12] In this case, there would actually never have been an unexpressed (or imperfectly expressed) meaning which can be captured by someone else. Instead, there would only have been the *impression* of having meant exactly what somebody else said, where this impression is based on a memory falsification.[13] As I have already said, I cannot currently rule out this possibility, nor can I cite any additional evidence which would help to strengthen the claim that said experiences are authentic and not based on false memory. But if we take seriously the impression of having meant exactly what someone else said without having been able to adequately express the content in question on one's own – as I think we should – then this provides good reason to unreservedly talk about unexpressed meanings even though we cannot provide clear-cut identity conditions for these entities.

[12] I follow the common account of déjà vu, according to which the experienced familiarity in a déjà vu is always illusory.

[13] It might not be entirely clear whether this situation should correctly be described as a case of déjà vu. According to the currently accepted standard definition, a déjà vu is "any subjectively inappropriate impression of familiarity of a present experience with an undefined past." (Brown, 2004, pp. 12 and 17) According to this definition, the situation presented (i.e., the student-professor case) cannot count as a déjà vu because there is, crucially, no subjective impression that the feeling of familiarity is inappropriate. On the contrary, the example only works if the impression of remembering and recognizing the content to be expressed is not perceived as being inappropriate. However, "[t]he amorphous nature of the experience [– namely the déjà vu experience –] is in part responsible for the difficulty in settling on a specific label, and the varied manner in which the experience is defined reflects this problem." (Brown, 2004, p. 17) So, there plausibly are other viable definitions of "déjà vu" – different from the standard definition cited before – which unproblematically cover the student-professor case because a subjective impression of inappropriateness or incorrectness is not included in the definition of what makes a déjà vu. The student-professor situation should then be described as the (illusory) feeling of having already thought what the utterance one hears (or reads) right now for the first time means. I think that this description warrants considering the student-professor situation as a case of déjà vu if the perceived familiarity, i.e., the recognition of what one thought in the meaning of someone else's words, is in fact illusory.

10.3 Unexpressed Content and the Conduit Metaphor

A final point needs to be raised following my defense of the legitimacy of including unexpressed, merely mental meaning in the discussion: Does a commitment to unexpressed meanings not cast the position defended here back to the conduit metaphor after I explicitly rejected it in Part I of this investigation? I claim that this is not the case. Accepting unexpressed meaning does not amount to a commitment to the conduit metaphor for the following three reasons.

First of all, the results from Part I of the present investigation stay in place. That there is unexpressed meaning does not cast any doubt on the claim that reasoning is constitutively dependent on language. This proves that thought (at least not all of thought) cannot be conceived of as being primary or prior to and independent of language. This result is in conflict with the conduit metaphor but compatible with the existence of unexpressed meaning.

10.3.1 The Conduit Metaphor Generalized

Secondly, the conduit metaphor concerns not only the relation between content and language, but also the relation between any means of expression and content. According to the conduit metaphor, language and content might stand in a relation analogous to a pipe and the water it conducts or to box and what it contains. According to this approach, the water (content) is generally independent from the conduit (language) just as the content of a box is independent from the container which carries the content. More importantly still, the pipe or container metaphor is not restricted to language. Any means of depiction, communication, or expression can be metaphorically equated with the conduit side of the conduit metaphor.

This approach inevitably invites us to think of content in terms of something which is akin to Kant's thing in itself. Thinking of content as being independent of any way whatsoever in which it can be presented – indepen-

dently of each and every format, we might say – is analogous to thinking of a thing independently of any way in which it can be perceived. Such a perspective on content leads to absurdity. It is, of course, legitimate to consider content in abstraction from any *particular* way in which it can be presented. This is crucial since, otherwise, we could not even consider the question of whether different ways of presentation (e.g., different expressions) might give expression to one and the same meaning, *viz.*, whether different presentations might be equivalent in terms of the content they give expression to. For example, when we consider whether "It is raining" is an adequate translation of "Es regnet," we naturally assume that there is something these sentences mean, and, in order to decide whether they mean the same, we need to consider their content in abstraction from their presentation: Here is the sentence, and here is what it means; and the sentence and what it means are different things. In the end, the sentence could have meant something else.

Abstracting away from any particular means of presentation in order to keep meaning constant across different ways of expressing one and the same content is legitimate and indispensable. But abstracting away from any particular way of presenting content is not the same as abstracting away from *every* way of presenting said content. The latter would leave us with content independently of every possible way of expressing it. This notion of content or meaning in itself, which is suggested by the conduit metaphor, is objectionable and can be compared, as I said, with Kant's notion of a thing in itself.[14] The notion of content as being abstracted away from any particular

[14] The explanatory analogy between Kant's thing in itself and content in itself should be confined to the explicitly mentioned aspect: Thinking of an object independently of whichever way it can be perceived is just as problematic as thinking of meaning independently of every possible form of presentation. I do not wish, however, to draw further parallels between the notions of the thing in itself and content in itself. Especially exegetical questions about the ontological status of the thing in itself can, I hope, be evaded at this point. It might be that content in itself is impossible, or it might be that content in itself does indeed exist but cannot be grasped because meaning can only be accessed in some form of presentation. I will not take a stance on these issues here since they do not seem to be crucial for the point in question. I should add, nevertheless, that the analogy between meaning in itself and Kant's thing in itself probably works best on the basis of an *one world, epistemic two aspects* interpretation of Kant's philosophy (cf. Heidemann, 2021, p. 3238), which is my preferred reading of Kant anyway.

way of expression, but not from every way of presentation, is, in contrast, a theoretical necessity and not objectionable in any way.

By accepting unexpressed meaning as a legitimate notion for enquiry, only the latter (unobjectionable and theoretically necessary) option is admitted. Even if content is conceived of as being unexpressed and merely mental, it is still conceived of as being formatted in some way – e.g., in a language of thought, pictorially, or as being presented in a mental map, etc. We can abstract away from any particular way of presentation, but we cannot abstract away from every way of presentation since this would amount to a notion of content without any presentation at all.[15] A content without any form of presentation at all (in contrast to a content conceived of independently from any presentation in particular) is indeed a figment of the philosopher.

10.3.2 Language as a Burden?

This leads to the third, related point: A consequence of this way of thinking of meaning in itself, motivated by the conduit metaphor, is a certain way of feeling that language (or any other way of presentation) somehow "blocks" a direct grasp of the content as it is in itself, i.e., independently of every possible representation. Linguistic expression is from this perspective conceived of as something which does not provide access to content but, rather, as something which disguises or veils the meaning we strive to grasp. As a proponent of this wrongheaded view, according to which meaning must be stripped of its (linguistic) clothing in order to be perceived in its purest possible form, we can cite Gottlob Frege, the grandfather of analytical philosophy (cf. Dummett, 2014, pp. 13 and 24).

"Frege held that we human beings have access to thoughts only as ex-

[15] Another helpful analogy, apart from Kant's thing in itself, might be Aristotle's hylomorphism: Everything is a compound of matter *and* form. The same matter (or content) might appear in different forms (or formats), but matter entirely without any form is impossible. (That the Aristotelian conception also works in the other direction, since we can also have the same form manifested in different matter, is not necessarily a downside of the analogy. We might also admit of the same vehicle – be it a picture, symbol, word, or what have you – with different meanings. The crucial aspect remains intact, namely that we cannot have a form devoid of any matter. However, the analogy is certainly not perfect and should not be stretched beyond its merely illustrative use.)

pressed in language or symbolism. He conceived of thoughts as intrinsically apt for linguistic expression[,]" Dummett (2014, p. 11) explains, but "[...] it was for him no contradiction to suppose beings who grasp in their nakedness, that is, without linguistic clothing, the same thoughts as we do." (Dummett, 2014, p. 11) What Dummett (2014, p. 11) calls grasping thoughts "[...] in their nakedness, that is, without linguistic clothing [...]" is what I call conceiving of a content in itself. Assuming that there is such content in itself is, as pointed out before, already a confusion. Although we can conceive of a thought without its "linguistic clothing" by abstracting away from a particular form of presentation, we cannot grasp thoughts "in their nakedness," i.e., as they are in themselves and independently of every form of presentation.

More importantly though, Frege lamented the fact that our access to thoughts is limited to linguistically (or in any other way representationally) mediated grasp of content. For him, "[t]he main task of the logician consists in *liberation from language* [...]" (Mohanty et al., 1974, p. 89; emphasis added),[16] and Frege complains:

> I am not in the happy position here of a mineralogist who shows his hearers a mountain crystal. I cannot put a thought in the hands of my readers with the request that they should minutely examine it from all sides. I have to content myself with presenting the reader with a thought, *in itself immaterial, dressed in sensible linguistic form.* The metaphorical aspect of language presents difficulties. The sensible always breaks in and makes expression metaphorical and so improper. So *a battle with language takes place* and I am compelled to occupy myself with language although it is not my proper concern here. (Frege, 1956, p. 298, n. 1; emphasis added)[17]

[16] This passage is also quoted in Dummett (2014, p. 7) and appears in Frege's letter to Husserl from October 30–November 1, 1906.

[17] In the German original it says:

> Ich bin hier nicht in der glücklichen Lage eines Mineralogen, der seinen Zuhörern einen Bergkristall zeigt. Ich kann meinen Lesern nicht einen Gedanken in die Hände geben mit der Bitte, ihn von allen Seiten recht genau zu betrachten. Ich muß mich begnügen, *den an sich unsinnlichen Gedanken in*

Certainly language can lead us astray in logical analysis,[18] but Frege sometimes seems to evince an almost hostile attitude towards language. This hostile attitude is, at times, even clearer in Wittgenstein: for example when he says that "Philosophy is a struggle against the bewitchment of our understanding by the resources of our language." (Wittgenstein, 2009, § 109)[19]

10.3.3 How Not to Be a Light Dove

These few remarks should be sufficient to prove that a presupposed conduit metaphor is a widespread phenomenon, even among the most astute philosophers. Trabant (cf. 2008, pp. 64-68) even suggests that seeing language through the lens of the conduit metaphor has been deeply engrained in Western philosophy at least since Plato and Aristotle. The negative effects such a presupposition can have are evident since it leads to a stark misjudgment regarding language: Namely that language was not so much an aid in getting access to content but, rather, an impediment to be overcome in order to reach content in itself. Yet, there is no content in itself, *viz.*, no meaning independently of every possible encoding, framing, presenting, formatting, or what have you. To think otherwise would be to make the same mistake as Kant's light dove:

> The light dove, in free flight cutting through the air the resistance of which it feels, could get the idea that it could do even better in airless space. (Kant, 1998a, p. 129 = A5/B8 f)[20]

die sinnliche sprachliche Form gehüllt dem Leser darzubieten. Dabei macht die Bildlichkeit der Sprache Schwierigkeiten. Das Sinnliche drängt sich immer wieder ein und macht den Ausdruck bildlich und damit uneigentlich. So entsteht *ein Kampf mit der Sprache* und ich werde genötigt, mich noch mit der Sprache zu befassen, obwohl das ja hier nicht meine eigentliche Aufgabe ist. (Frege, 2010, pp. 97 f, n. 4 [p. 66]; emphasis added)

[18] A case in point was presented and discussed on page 88.

[19] "Die Philosophie ist ein Kampf gegen die Verhexung unsres Verstandes durch die Mittel unserer Sprache." (Wittgenstein, 2003b, § 109)

[20] The passage occurs in the context of a criticism of Platonic metaphysics and goes as follows in the German original:

> Die leichte Taube, indem sie im freien Fluge die Luft teilt, deren Widerstand

In this analogy, the philosopher who follows the conduit metaphor and therefore comes to conceive of language as an obstacle for his reaching out to "pure" or "naked" content (in itself) thinks – just like the dove – that he might be able to advance his intellectual attempts by overcoming the very medium which makes his entire venture possible in the first place. Just like in the case of the dove, realizing this ambition would get our philosopher nowhere, much less off the ground. Language is an indispensable medium and instrument for dealing with content. It might not be fool-proof, but it is by far the best we could hope for. And, importantly, nothing – in principle – prevents us from perpetually enhancing it further and further.

10.4 Conclusion

After the conduit metaphor was ruled out in Part I (since reasoning, which is an important part of thought, is constitutively dependent on language), we returned to this view on the relation between language and thought again to highlight that it is not only false but can also prove to be very misleading. However, no deeper importance should be read into the cyclic path taken in the investigation at hand. Had I taken care of Binkley's critique (as it can be directed against the use I make of the Principle of Expressibility, despite the fact that it is ineffective against Searle's methodological use of the principle) earlier instead of procrastinating until the very last chapter, the thesis at hand could also have finished with the defense of the Principle of Expressibility, culminating in chapter 9.

> sie fühlt, könnte die Vorstellung fassen, daß es ihr im luftleeren Raum noch viel besser gelingen werde. (Kant, 1998b, pp. 52-55 = A5/B8 f)

Tarbet (1968, p. 259) thinks that "Kant's choice of the beautiful dove, a symbol of peace and love, to represent reason's metaphysical soaring is revealing. If he was as opposed to metaphysics as some suggest, why did he not use a less attractive bird?" However, no such hidden sympathy for the conduit metaphor should be read into my use of the quote here. I make use of Kant's metaphor because I find it beautiful and apt for the present purpose. Also, I find Tarbet's suggestion quite flimsy since this kind of speculation "[...] to make inferences concerning Kant's character [...]" (Tarbet, 1968, p. 263) because "[...] metaphors present a view of the philosopher as well as of the philosophy" (Tarbet, 1968, p. 270) is to be considered as highly dubious in any case and has no legitimate place in a proper philosophical investigation, according to my opinion.

In this final Part III, I argued that the correct relation between what can be thought (set M) and what can be expressed (set S) is option 3 – what can be meant being a proper subset of what can be said – as introduced on page 136 and depicted on page 140. This was achieved by ruling out competing options and arguing that we have no reason to assume that what can be thought might transcend the domain of what can be said by defusing apparently pressing problem cases for the Principle of Expressibility. Dealing with the scope of the domains of the "sayable" and the "meanable," Part III stands orthogonal to the topics covered in Parts I and II, thereby presenting an entirely different perspective on the relation between mind and language.

Part II is concerned with ruling out linguistic relativity as a popular account of how language and mind relate to each other. This was done, however, not by providing a direct argument against linguistic relativity but by undercutting what was identified as an important underlying assumption, shared by many linguistic relativists, which (presumably) motivates adoption of a linguistic relativity theory in the first place in many cases. Although the demonstration that language and culture are not inseparably intertwined does not refute linguistic relativism *per se*, it might nevertheless diminish its attractiveness considerably if the widely held assumption about the inextricableness of language and culture should turn out to be a major factor in lending credibility to linguistic relativity, as I speculatively presume it does for many scholars.

The conduit metaphor can be considered as the converse extreme of linguistic relativity. Therefore, Parts I and II are plausibly much closer related to each other than to Part III. The uniting factor is that Parts I and II are both concerned with the question of which effect language and mind have on each other whereas Part III takes an entirely independent perspective, not concerned with how thought and language affect each other, but with questions of the scope of language and mind respectively. These quite different questions – concerning effect vs. scope – can nevertheless legitimately be classified as fundamental questions regarding the relation between mind and language. The present investigation can therefore said to approach a single topic – how mind and language relate to each other – from different perspec-

tives to achieve a fuller picture of the intricate and multi-faceted relation between language and thought.

References

Abbott, E. A. (2010). *Flatland: An Edition with Notes and Commentary*. Cambridge and Washington, D.C: Cambridge University Press and Mathematical Association of America. https://doi.org/10.1017/CBO9781139194921.

Adelson, E. H. (1995). Checkershadow Illusion. http://persci.mit.edu/gallery/checkershadow.

Ahearn, L. M. (2017). *Living Language: An Introduction to Linguistic Anthropology*. Number 2 in Primers in Anthropology. Malden, Mass.: Wiley-Blackwell, second edition.

Andrews, K. & Monsó, S. (2021). Animal Cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition. https://plato.stanford.edu/archives/spr2021/entries/cognition-animal/.

Anscombe, G. E. M. (1963). *Intention*. Cambridge, Mass. & London, UK: Harvard University Press, second edition.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Austin, J. L. (1970). Performative Utterances. In J. O. Urmson & G. J. Warnock (Eds.), *Philosophical papers* (pp. 233–252). Oxford: Clarendon Press, second edition.

Ayer, A. J. & Rhees, R. (1954). Can There Be a Private Language? *Aristotelian Society Supplementary Volume*, 28(1), 63–94. https://doi.org/10.1093/aristoteliansupp/28.1.63.

Baker, G. (1998). The private language argument. *Language & Communication*, 18(4), 325–356. https://doi.org/10.1016/S0271-5309(98)

00010-X.

Bannow, T. (2009). Local company creates Klingon dictionary. *The Minnesota Daily*. https://mndaily.com/186847/uncategorized/local-company-creates-klingon-dictionary/.

Bartlett, P. (2009). Artificial Languages. In K. Brown & S. Ogilvie (Eds.), *Concise Encyclopedia of Languages of the World* (pp. 75–78). Amsterdam: Elsevier, 1st edition.

Bayne, T. & Montague, M., Eds. (2011a). *Cognitive Phenomenology*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199579938.001.0001.

Bayne, T. & Montague, M. (2011b). Cognitive Phenomenology: An Introduction. In T. Bayne & M. Montague (Eds.), *Cognitive Phenomenology* (pp. 1–34). Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199579938.003.0001.

Beaney, M. (2010). Verificationism. In A. Barber, R. J. Stainton, & K. Brown (Eds.), *Concise Encyclopedia of Philosophy of Language and Linguistics* (pp. 810–813). Amsterdam: Elsevier.

Benacerraf, P. (1973). Mathematical Truth. *The Journal of Philosophy*, 70(19), 661–679. https://doi.org/10.2307/2025075.

Bergen, B. K. (2001). Nativization processes in L1 Esperanto. *Journal of Child Language*, 28(03). https://doi.org/10.1017/s0305000901004779.

Berkeley, G. (1999a). Principles of Human Knowledge. In H. Robinson (Ed.), *Principles of Human Knowledge and Three Dialogues*, Oxford World's Classics (pp. 1–95). Oxford: Oxford University Press, paperback edition.

Berkeley, G. (1999b). *Principles of Human Knowledge and Three Dialogues*. Oxford World's Classics. Oxford: Oxford University Press, paperback edition.

Berlin, I. (1939). Verification. *Proceedings of the Aristotelian Society*, 39(1), 225–248. https://doi.org/10.1093/aristotelian/39.1.225.

Bermúdez, J. L. (2003). *Thinking without Words*. Philosophy of Mind Series. New York: Oxford University Press.

Berto, F. & Nolan, D. (2021). Hyperintensionality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition. https://plato.stanford.edu/archives/spr2021/entries/hyperintensionality/.

Bertolet, R. (1999). Private Language Argument. In R. Audi (Ed.), *The Cambridge Dictionary of Philosophy* (pp. 741). Cambridge and New York: Cambridge University Press, second edition.

Bickerton, D. (1996). *Language and Human Behavior*. The Jessie and John Danz Lectures. Seattle: University of Washington Press, 2nd print edition.

Binkley, T. (1979). The Principle of Expressibility. *Philosophy and Phenomenological Research*, 39(3), 307–325. https://doi.org/10.2307/2106489.

Block, N. J. (2007). Mental Paint. In *Consciousness, Function, and Representation*, volume 1 of *Collected Papers* (pp. 533–570). Cambridge, Mass.: MIT Press.

Blumberg, K. & Holguín, B. (2018). Ultra-liberal attitude reports. *Philosophical Studies*, 175(8), 2043–2062. https://doi.org/10.1007/s11098-017-0949-7.

Bodmer, F. (1946). *The Loom of Language: A Guide to Foreign Languages for the Home Student*. Number 3 in Primers for the Age of Plenty. London: Allen & Unwin, 4th impression edition.

Boghossian, P. (2014). What is inference? *Philosophical Studies*, 169(1), 1–18. https://doi.org/10.1007/s11098-012-9903-x.

Boghossian, P. (2016). Rationality, reasoning and rules: Reflections on Broome's rationality through reasoning. *Philosophical Studies*, 173(12), 3385–3397. https://doi.org/10.1007/s11098-016-0716-1.

BonJour, L. (1991). Is thought a symbolic process? *Synthese*, 89(3), 331–352. https://doi.org/10.1007/BF00413501.

BonJour, L. (1992). Analytic Philosophy and the Nature of Thought. (Unpublished manuscript). http://faculty.washington.edu/bonjour/Unpublished%20articles/UBCPAPER.html.

Boyle, M. (2012). Essentially Rational Animals. In G. Abel & J. Co-

nant (Eds.), *Rethinking Epistemology. Volume 2* (pp. 395–427). Berlin and Boston: De Gruyter. https://doi.org/10.1515/9783110277944.395.

Boyle, M. (2016). Additive Theories of Rationality: A Critique. *European Journal of Philosophy*, 24(3), 527–555. https://doi.org/10.1111/ejop.12135.

Braddon-Mitchell, D. & Jackson, F. (2007). *The Philosophy of Mind and Cognition*. Malden, Mass.: Blackwell Publishing, second edition.

Brown, A. S. (2004). *The Déjà vu Experience*. Essays in Cognitive Psychology. New York and Hove: Psychology Press.

Brown, A. S. (2012). *The Tip of the Tongue State*. Essays in Cognitive Psychology. New York and London: Psychology Press.

Brown, J. R. (2011). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. Philosophical Issues in Science. New York: Routledge, second edition.

Camp, E. (2007). Thinking With Maps. *Philosophical Perspectives*, 21(1), 145–182. https://doi.org/10.1111/j.1520-8583.2007.00124.x.

Carnap, R. (1931). Überwindung der Metaphysik durch logische Analyse der Sprache. *Erkenntnis*, 2(1), 219–241. https://doi.org/10.1007/BF02028153.

Carnap, R. (1966). The Elimination of Metaphysics Through Logical Analysis of Language. In A. J. Ayer (Ed.), *Logical Positivism*, The Library of Philosophical Movements (pp. 60–81). New York: The Free Press, 2nd printing edition.

Carroll, L. (1895). What the Tortoise said to Achilles. *Mind*, 4(14), 278–280. https://doi.org/10.1093/mind/IV.14.278.

Carruthers, P. (2002). The cognitive functions of language. *The Behavioral and Brain Sciences*, 25(6), 657–674; discussion 674–725. https://doi.org/10.1017/S0140525X02000122.

Carruthers, P. & Boucher, J. (1998). Introduction: Opening up options. In P. Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary Themes* (pp. 1–18). Cambridge: Cambridge University Press, 1st edition. https://doi.org/10.1017/CBO9780511597909.002.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: M.I.T. Press.

Chudnoff, E. (2015). *Cognitive Phenomenology*. New Problems of Philosophy. London and New York: Routledge. https://doi.org/10.4324/9781315771922.

Coles, A. (2018). Teacher's shock after learning boy, 4, who barely talked speaks fluent Klingon. *The Mirror - Onlie edition - mirror.co.uk*. https://www.mirror.co.uk/news/us-news/nursery-worker-stunned-learn-boy-13304918.

Conee, E. (1994). Phenomenal knowledge. *Australasian Journal of Philosophy*, 72(2), 136–150. https://doi.org/10.1080/00048409412345971.

Corsetti, R. (1996). A Mother Tongue Spoken Mainly by Fathers. *Language Problems and Language Planning*, 20(3), 263–273. https://doi.org/10.1075/lplp.20.3.05cor.

David, M. (2016). The Correspondence Theory of Truth. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2016 edition. https://plato.stanford.edu/archives/fall2016/entries/truth-correspondence/.

Davidson, D. (1991). Three Varieties of Knowledge. *Royal Institute of Philosophy Supplement*, 30, 153–166. https://doi.org/10.1017/S1358246100007748.

Davidson, D. (2001). Rational Animals. In *Subjective, Intersubjective, Objective* (pp. 95–105). Oxford: Clarendon Press, reprinted 2004 edition. https://doi.org/10.1093/0198237537.003.0007.

DePoe, J. M. (2011). Berkeley's Master Argument for Idealism. In M. Bruce & S. Barbone (Eds.), *Just the Arguments* (pp. 68–69). Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444344431.ch16.

Descartes, R. (1982). *Principles of Philosophy*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-7888-1.

Descartes, R. (2008). *Meditations on First Philosophy: With Selections from the Objections and Replies*. Oxford: Oxford University Press.

Devitt, M. (2006). *Ignorance of Language*. New York: Oxford University Press. https://doi.org/10.1093/0199250960.001.0001.

Donnellan, K. S. (1966). Reference and Definite Descriptions. *The Philosophical Review*, 75(3), 281–304. https://doi.org/10.2307/2183143.

Donnellan, K. S. (2012). Kripke and Putnam on Natural Kind Terms. In J. Almog & P. Leonardi (Eds.), *Essays on Reference, Language, and Mind* (pp. 178–203). Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199857999.003.0007.

Du Bois Reymond, E. (1884). *Über die Grenzen des Naturerkennens: Die sieben Welträthsel. Zwei Vorträge*. Leipzig: De Gruyter, reprint 2020, 2nd edition. https://doi.org/10.1515/9783112375969.

Dummett, M. (2014). *Origins of Analytical Philosophy*. London and New York: Bloomsbury.

Eco, U. (1995). *The Search for the Perfect Language*. The Making of Europe. Oxford, UK and Cambridge, Mass., USA: Blackwell.

Enfield, N. J., Kockelman, P., & Sidnell, J. (2014). Directions in the anthropology of language. In N. Enfield, P. Kockelman, & J. Sidnell (Eds.), *The Cambridge Handbook of Linguistic Anthropology* (pp. 1–24). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139342872.001.

Fine, K. (1994). Essence and Modality. *Philosophical Perspectives*, 8, 1–16. https://doi.org/10.2307/2214160.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199548774.001.0001.

Fraissler, H. (2014). *Identität bei Saul Kripke. Oder: Was uns Schmidentität über Identität sagen kann / Saul Kripke on Identity. Or: What Schmidentity can tell us about Identity*. Diploma Thesis, Karl-Franzens-Universität Graz, Graz. http://media.obvsg.at/p-AC12144087-2001.

Fraissler, H. (2021). A Private Language Argument to elucidate the relation between mind and language. *Filosofia Unisinos*, 22(1), 48–58. https://doi.org/10.4013/fsu.2021.221.06.

Frankish, K. & Evans, J. S. B. T. (2009). The duality of mind: An histor-

ical perspective. In J. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 1–29). Oxford: Oxford University Press, reprinted 2013 edition. https://doi.org/10.1093/acprof:oso/9780199230167.003.0001.

Frege, G. (1956). The Thought: A Logical Inquiry. *Mind*, 65(1), 289–311. https://doi.org/10.1093/mind/65.1.289.

Frege, G. (1960). On Sense and Reference. In M. Black & P. Geach (Eds.), *Translations from the Philosophical Writings of Gottlob Frege* (pp. 56–78). Oxford: Basil Blackwell, second edition.

Frege, G. (2010). Der Gedanke. Eine logische Untersuchung. In W. Künne (Ed.), *Die Philosophische Logik Gottlob Freges: Ein Kommentar mit den Texten des Vorworts zu Grundgesetze der Arithmetik und der Logischen Untersuchungen I-IV*, number 30 in Klostermann RoteReihe (pp. 87–112 [58–77]). Frankfurt am Main: Vittorio Klostermann.

Gallois, A. (1974). Berkeley's Master Argument. *The Philosophical Review*, 83(1), 55. https://doi.org/10.2307/2183873.

Girle, R. (2009). *Modal Logics and Philosophy*. Durham: Acumen, second edition. https://doi.org/10.1017/UPO9781844654536.

Grice, H. P. (1991). Logic and Conversation. In *Studies in the Way of Words* (pp. 22–40). Cambridge, Mass. & London, UK: Harvard University Press, first harvard university press paperback edition.

Gumperz, J. J. & Levinson, S. C. (1996). Linguistic Relativity Re-Examined. In J. Gumperz & S. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 1–18). Cambridge: Cambridge University Press.

Hackstette, K. (1982). On Searle's Principle of Expressibility. *Studies in Language*, 6(3), 425–430. https://doi.org/10.1075/sl.6.3.08hac.

Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4, 31–52. https://doi.org/10.2307/2214186.

Heck, R. G. (2000). Nonconceptual Content and the "Space of Reasons". *Philosophical Review*, 109(4), 483–523. https://doi.org/10.1215/00318108-109-4-483.

Heidegger, M. (1949). What is Metaphysics? In *Existence and Being* (pp. 355–392). Chicago: H. Regnery.

Heidegger, M. (1955). *Was ist Metaphysik?* Frankfurt am Main: Vittorio Klostermann, seventh edition.

Heidemann, D. (2021). Kant and the forms of realism. *Synthese*, 198(S13), 3231–3252. https://doi.org/10.1007/s11229-019-02502-4.

Heyes, C. M. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Hiskey, D. (2012). A Man Once Tried to Raise His Son as a Native Speaker in Klingon. *Today I Found Out*. http://www.todayifoundout.com/index.php/2012/08/a-man-once-tried-to-raise-his-son-as-a-native-speaker-in-klingon/.

Hockett, C. F. (1976). The Problem of Universals in Language. In J. H. Greenberg (Ed.), *Universals of Language: Report of a Conference Held at Dobbs Ferry, New York, April 13-15, 1961*, number 37 in The M.I.T. Press Paperback Series (pp. 1–29). Cambridge, Mass.: M.I.T. Press, 5th printing, 2nd edition.

Hofmann, F. (2021). Explaining free will by rational abilities. (Unpublished manuscript).

Holden, T. (2019). Berkeley on Inconceivability and Impossibility. *Philosophy and Phenomenological Research*, 98(1), 107–122. https://doi.org/10.1111/phpr.12432.

Holyoak, K. J. & Morrison, R. G. (2005). Thinking and Reasoning: A Reader's Guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 1–9). Cambridge: Cambridge University Press.

Holyoak, K. J. & Morrison, R. G. (2013). Thinking and Reasoning: A Reader's Guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning*, Oxford Library of Psychology (pp. 1–7). Oxford and New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0001.

Horn, L. (1997). Presupposition and Implicature. In S. Lappin (Ed.), *The Handbook of Contemporary Semantic Theory*, number 3 in Blackwell Handbooks in Linguistics (pp. 299–319). Oxford, UK and Malden, Mass.: Blackwell Publishers, paperback edition.

Horn, L. R. (2006). Implicature. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 2–28). Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756959.ch1.

Horowitz, A. (2011). Jackson's Knowledge Argument. In M. D. Bruce & S. Barbone (Eds.), *Just the Arguments: 100 of the Most Important Arguments in Western Philosophy* (pp. 320–323). Chichester, West Sussex, U.K: Wiley-Blackwell. https://doi.org/10.1002/9781444344431.ch84.

Hume, D. (2007). *An Enquiry Concerning Human Understanding*. Oxford World's Classics. Oxford and New York: Oxford University Press.

Hurley, S. & Nudds, M. (2006). The questions of animal rationality: Theory and evidence. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 1–84). Oxford and New York: Oxford University Press, reprinted 2010 edition. https://doi.org/10.1093/acprof:oso/9780198528272.003.0001.

Iacobini, C. (2006). Morphological Typology. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 278–282). Amsterdam and Boston: Elsevier, second edition. https://doi.org/10.1016/B0-08-044854-2/00155-3.

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, 32(127), 127–136. https://doi.org/10.2307/2960077.

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy*, 83(5), 291–295. https://doi.org/10.2307/2026143.

Jones, J.-E. (2018). Locke on Real Essence. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition. https://plato.stanford.edu/archives/fall2018/entries/real-essence/.

Kahneman, D. (2013). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 1st paperback edition.

Kannetzky, F. (2001). The Principle of Expressibility and Private Language. *Acta Philosophica Fennica*, 69, 191–212. http://www.musikwissenschaft.uni-bremen.de/fileadmin/redak_philo/Papers/Kannetzky/Kannetzky_-_The_Principle_of

_Expressibility_and_Private_Language.pdf.

Kannetzky, F. (2002). Expressibility, Explicability, and Taxonomy. Some Remarks on the Principle of Expressibility. In G. Grewendorf & G. Meggle (Eds.), *Speech Acts, Mind, and Social Reality*, volume 79 of *Studies in Linguistics and Philosophy* (pp. 65–82). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-0589-0_5.

Kant, I. (1998a). *Critique of Pure Reason*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge and New York: Cambridge University Press.

Kant, I. (1998b). *Kritik der reinen Vernunft*. Number 505 in Philosophische Bibliothek. Hamburg: Felix Meiner Verlag. https://doi.org/10.28937/978-3-7873-2112-4.

Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481–563). New York and Oxford: Oxford University Press.

King, J. C. (2017). Structured Propositions. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University, fall 2017 edition. https://plato.stanford.edu/archives/fall2017/entries/propositions-structured/.

von Kleist, H. (1924). Über die allmähliche Verfertigung der Gedanken beim Reden. In *Heinrich von Kleist. Sämtliche Werke* (pp. 817–822). Vienna: Phaidon Verlag.

von Kleist, H. (2004a). *Heinrich von Kleist. Selected Writings*. Indianapolis/Cambridge: Hackett Publishing Company, reprinted edition.

von Kleist, H. (2004b). On the Gradual Production of Thoughts Whilst Speaking. In D. Constantine (Ed.), *Heinrich von Kleist. Selected Writings* (pp. 405–409). Indianapolis/Cambridge: Hackett Publishing Company, reprinted edition.

Kramsch, C. (1998). *Language and Culture*. Oxford Introductions to Language Study. Oxford: Oxford University Press.

Kramsch, C. (2004). Language, Thought, and Culture. In A. Davies & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 235–261).

Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757000.ch9.

Krause, E.-D. (1999). *Großes Wörterbuch Esperanto - Deutsch*. Hamburg: Buske.

Kripke, S. A. (1981). *Naming and Necessity*. Oxford, UK and Cambridge, USA: Blackwell Publishing, 23rd paperback, 2012 edition.

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, Mass.: Harvard University Press, eighth printing, 1995 edition.

Kripke, S. A. (2011a). Identity and Necessity. In *Philosophical Troubles. Collected Papers, Volume 1* (pp. 1–26). Oxford and New York: Oxford University Press. https://www.doi.org/10.1093/acprof:oso/9780199730155.003.0001.

Kripke, S. A. (2011b). Unrestricted Exportation and Some Morals for the Philosophy of Language. In *Philosophical Troubles. Collected Papers, Volume 1* (pp. 322–350). Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199730155.003.0011.

Kripke, S. A. (2013). *Reference and Existence: The John Locke Lectures*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199928385.001.0001.

Kukla, A. (2005). *Ineffability and Philosophy*. Number 22 in Routledge Studies in Twentieth-Century Philosophy. London and New York: Routledge. https://doi.org/10.4324/9780203325070.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202–251). Cambridge: Cambridge University Press, second edition. https://doi.org/10.1017/CBO9781139173865.013.

Laland, K. N. & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology: Issues, News, and Reviews*, 12(3), 150–159. https://doi.org/10.1002/evan.10111.

Laland, K. N. & Janik, V. M. (2006). The animal cultures debate. *Trends in Ecology & Evolution*, 21(10), 542–547. https://doi.org/10.1016/

j.tree.2006.06.005.

Leibniz, G. W. (1989). The Principles of Philosophy, or, the Monadology. In R. Ariew & D. Garber (Eds.), *G.W. Leibniz. Philosophical Essays* (pp. 213–225). Indianapolis and Cambridge: Hackett Publishing Company.

Lewis, D. (1983). Languages and Language. In *Philosophical Papers. Volume I* (pp. 163–188). New York and Oxford: Oxford University Press. https://doi.org/10.1093/0195032047.003.0011.

Lewis, D. K. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Lewis, D. K. (1999). What experience teaches. In *Papers in Metaphysics and Epistemology*, volume 2 of *Cambridge Studies in Philosophy* (pp. 262–290). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511625343.018.

Lightner, D. T. (1997). Hume on Conceivability and Inconceivability. *Hume Studies*, 23(1), 113–132. https://doi.org/10.1353/hms.2011.0127.

Locke, J. (1975). *An Essay Concerning Human Understanding*. Oxford and New York: Oxford University Press, reprinted 2011 edition.

Löffler, W. (2004). "Esperanto. The Feeling of Disgust": Wittgenstein on Planned Languages. *From the ALWS archives: A selection of papers from the International Wittgenstein Symposia in Kirchberg am Wechsel*, (pp. 209–211). http://wittgensteinrepository.org/agora-alws/article/view/2529.

Lohmar, D. (2016). Non-Linguistic Thinking and Communication—Its Semantics and Some Applications. In T. Breyer & C. Gutland (Eds.), *Phenomenology of Thinking: Philosophical Investigations into the Character of Cognitive Experiences*, number 4 in Routledge Research in Phenomenology (pp. 165–182). New York: Routledge.

Look, B. C. (2018). Arguments for the Existence of God. In M. R. Antognazza (Ed.), *The Oxford Handbook of Leibniz* (pp. 701–716). Oxford and New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199744725.013.010.

Lucy, J. A. (1997). Linguistic Relativity. *Annual Review of Anthropology*, 26(1), 291–312. https://doi.org/10.1146/annurev.anthro.26.1.291.

Lurz, R. W. (2009). The philosophy of animal minds: An introduction. In R. W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 1–14). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819001.001.

Magidor, O. (2009). Category mistakes are meaningful. *Linguistics and Philosophy*, 32(6), 553–581. https://doi.org/10.1007/s10988-010-9067-0.

Magidor, O. (2013). *Category Mistakes*. Oxford Philosophical Monographs. Oxford: Oxford University Press, 1st edition.

Mair, C. (2015). *English Linguistics: An Introduction*. Narr Bachelor-Wissen.De. Tübingen: Narr Francke Attempto, 3rd updated edition.

Malcolm, N. (1954). Wittgenstein's Philosophical Investigations. *The Philosophical Review*, 63(4), 530–559. https://doi.org/10.2307/2182289.

Mayer, H. (1992). *Grundwortschatz Esperanto: ein Lernwörterbuch*. Number 9 in Serio Instruo. Vienna: Pro Esperanto.

McHugh, C. & Way, J. (2015). Broome on Reasoning. *Teorema: Revista Internacional de Filosofía*, 34(2), 131–140. http://www.jstor.org/stable/43694673.

McHugh, C. & Way, J. (2018). What is Reasoning? *Mind*, 127(505), 167–196. https://doi.org/10.1093/mind/fzw068.

McWhorter, J. H. (2014). *The Language Hoax: Why the World Looks the Same in Any Language*. Oxford: Oxford University Press.

Mercier, H. & Sperber, D. (2017). *The Enigma of Reason*. Cambridge, Mass.: Harvard University Press. https://doi.org/10.4159/9780674977860.

Mohanty, J. N., Frege, G., & Husserl, E. (1974). Frege-Husserl Correspondence. *Southwestern Journal of Philosophy*, 5(3), 83–95. https://doi.org/10.5840/swjphil19745338.

Morgan, C. L. (1896). *An Introduction to Comparative Psychology*. London: Walter Scott.

Morris, W. E. & Brown, C. R. (2021). David Hume. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition. https://plato.stanford

.edu/archives/spr2021/entries/hume/.

Navarro-Reyes, J. (2009). Can we say what we mean? Expressibility and background. *Pragmatics & Cognition*, 17(2), 283–308. https://doi.org/10.1075/pc.17.2.04nav.

Nida-Rümelin, M. (2010). Thinking without Language. A Phenomenological Argument for its Possibility and Existence. *Grazer Philosophische Studien*, 81(1), 55–75. https://doi.org/10.1163/9789042030190_005.

Okrand, M. (1992). *The Klingon Dictionary. English-Klingon, Klingon-English*. New York: Pocket Books.

Okrent, A. (2009). *In the Land of Invented Languages: Esperanto Rock Stars, Klingon Poets, Loglan Lovers, and the Mad Dreamers Who Tried to Build a Perfect Language*. New York: Spiegel & Grau, 1st edition.

Partee, B. H. (2007). Compositionality and Coercion in Semantics: The Dynamics of Adjective Meaning. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, number 190 in Verhandelingen / Koninklijke Nederlandse Akademie van Wetenschappen, Afd. Letterkunde (pp. 145–161). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Partee, B. H. (2010). Privative Adjectives: Subsective Plus Coercion. In R. Bäuerle, U. Reyle, T. Zimmermann, K. von Heusinger, & K. Turner (Eds.), *Presuppositions and Discourse: Essays Offered to Hans Kamp*, volume 21 of *Current research in the semantics/pragmatics interface* (pp. 273–285). Bingley: Emerald, 1st edition. https://doi.org/10.1163/9789004253162_011.

Preti, C. (2003). *On Kripke*. Wadsworth Philosophers Series. South Melbourne: Thomson/Wadsworth.

Prisco, J. (2018). How do you design a language from scratch? Ask a Klingon. *CNN Style*. https://www.cnn.com/style/article/star-trek-klingon-marc-okrand/index.html.

Putnam, H. (1975). The Meaning of "Meaning". *Minnesota Studies in the Philosophy of Science*, 7, 131–193.

Quine, W. V. O. (1948). On What There Is. *The Review of Metaphysics*, 2(5), 21–38. https://www.jstor.org/stable/20123117.

Quine, W. V. O. (1951a). Ontology and Ideology. *Philosophical Studies*, 2(1), 11–15. https://doi.org/10.1007/BF02198233.

Quine, W. V. O. (1951b). Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), 20–43. https://doi.org/10.2307/2181906.

Quine, W. V. O. (1969). Speaking of Objects. In *Ontological Relativity and Other Essays*, number One in The John Dewey Essays in Philosophy (pp. 1–25). New York: Columbia University Press. https://doi.org/10.7312/quin92204-002.

Raleigh, T. (2019). Wittgenstein, Spatial Phenomenology, and "The Private Language Argument". In T. Cheng, O. Deroy, & C. Spence (Eds.), *Spatial Senses. Philosophy of Perception in an Age of Science*, number 122 in Routledge Studies in Contemporary Philosophy (pp. 70–91). New York: Routledge. https://doi.org/10.4324/9781315146935-5.

Rawls, J. (1955). Two Concepts of Rules. *The Philosophical Review*, 64(1), 3–32. https://doi.org/10.2307/2182230.

Recanati, F. (2003). The Limits of Expressibility. In B. Smith (Ed.), *John Searle* (pp. 189–213). Cambridge: Cambridge University Press, first edition. https://doi.org/10.1017/CBO9780511613999.009.

Reddy, M. J. (1993). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 164–201). Cambridge: Cambridge University Press, second edition. https://doi.org/10.1017/CBO9781139173865.012.

Reines, M. F. & Prinz, J. (2009). Reviving Whorf: The Return of Linguistic Relativity. *Philosophy Compass*, 4(6), 1022–1032. https://doi.org/10.1111/j.1747-9991.2009.00260.x.

Rescorla, M. (2009). Chrysippus' dog as a case study in non-linguistic cognition. In R. W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 52–71). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819001.004.

Rescorla, M. (2019). The Language of Thought Hypothesis. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition. https://plato.stanford.edu/archives/sum2019/entries/language-thought/.

Russell, B. (1911). Knowledge by Acquaintance and Knowledge by Description. *Proceedings of the Aristotelian Society*, 11(1), 108–128. https://doi.org/10.1093/aristotelian/11.1.108.

Russell, B. (1995). *An Inquiry into Meaning and Truth: The William James Lectures for 1940 Delivered at Harvard University.* London and New York: Routledge, revised edition.

Salmon, N. (1986). *Frege's Puzzle.* Cambridge, Mass.: MIT Press.

Schor, E. (2007). Crocodiling in Esperanto On the Streets of Hanoi. *The Forward.* https://forward.com/culture/11460/crocodiling-in-esperanto-on-the-streets-of-hanoi-00361/.

Searle, J. R. (1964). How to Derive "Ought" From "Is". *The Philosophical Review*, 73(1), 43–58. https://doi.org/10.2307/2183201.

Searle, J. R. (1979a). Introduction. In *Expression and Meaning: Studies in the Theory of Speech Acts* (pp. vii–xii). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609213.001.

Searle, J. R. (1979b). Metaphor. In *Expression and Meaning: Studies in the Theory of Speech Acts* (pp. 76–116). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609213.006.

Searle, J. R. (1979c). Referential and attributive. In *Expression and Meaning: Studies in the Theory of Speech Acts* (pp. 137–161). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609213.008.

Searle, J. R. (1979d). A taxonomy of illocutionary acts. In *Expression and Meaning: Studies in the Theory of Speech Acts* (pp. 1–29). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609213.003.

Searle, J. R. (1983). *Intentionality, an Essay in the Philosophy of Mind.* Cambridge and New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139173452.

Searle, J. R. (1995). *The Construction of Social Reality.* New York: Free Press.

Searle, J. R. (2011). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press, 34th print edition.

Searle, J. R. (2013). What is a Speech Act? In M. Black (Ed.), *Philosophy in America*, number V in Muirhead Library of Philosophy (pp. 221–239). London & New York: Routledge.

Shepard, R. N. (1990). *Mind Sights: Original Visual Illusions, Ambiguities, and Other Anomalies, with a Commentary on the Play of Mind in Perception and Art*. New York: W.H. Freeman and Co.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, 25(5), 701–721. https://doi.org/10.1023/A:1020867916902.

Stenmark, M. (2013). Scientism. In A. L. C. Runehov & L. Oviedo (Eds.), *Encyclopedia of Sciences and Religions* (pp. 2103–2105). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-8265-8_1534.

Stern, D. G. (2011). Private Language. In M. McGinn & O. Kuusela (Eds.), *The Oxford Handbook of Wittgenstein* (pp. 333–350). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199287505.003.0016.

Stockwell, P. (2006). Invented Language in Literature. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 3–10). Amsterdam and Boston: Elsevier, second edition. https://doi.org/10.1016/B0-08-044854-2/00519-8.

Swift, J. (2019). *Gulliver's Travels*. Richmond, Surrey, UK: Alma Classics.

Tarbet, D. W. (1968). The Fabric of Metaphor in Kant's Critique of Pure Reason. *Journal of the History of Philosophy*, 6(3), 257–270. https://doi.org/10.1353/hph.2008.1383.

Tegmark, M. (2015). *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. London: Penguin Books.

Thomson, J. J. (1964). Private Languages. *American Philosophical Quarterly*, 1(1), 20–31.

Torin, A. (2010). Mary the Color Scientist. In E. B. Goldstein (Ed.), *Encyclopedia of Perception* (pp. 545–547). Los Angeles: SAGE.

Trabant, J. (2008). *Was ist Sprache?* Number 1844 in Beck'sche Reihe. München: Beck, original edition.

Tye, M. (1996). *Ten Problems of Consciousness: A Representational Theory*

of the Phenomenal Mind. Representation and Mind Series. Cambridge, Mass.: MIT Press, second printing edition.

Tye, M. (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts*. Representation and Mind Series. Cambridge, Mass.: MIT Press.

udiprod (2012). A Visual Riddle (The Epitaph of Stevinus). https://www.youtube.com/watch?v=nDKGHGdXLEg.

Uebel, T. (2019). Verificationism and (Some of) its Discontents. *Journal for the History of Analytical Philosophy*, 7(4), 1–31. https://doi.org/10.15173/jhap.v7i4.3535.

Vaughan, R. C. (2016). Goldbach's Conjectures: A Historical Perspective. In J. F. Nash & M. T. Rassias (Eds.), *Open Problems in Mathematics* (pp. 479–520). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32162-2_16.

Voltolini, A. (2017). Varieties of Cognitive Phenomenology. *Phenomenology and Mind*, 10, 94–107. https://doi.org/10.13128/Phe_Mi-20094.

Wacewicz, S. & Żywiczyński, P. (2015). Language Evolution: Why Hockett's Design Features are a Non-Starter. *Biosemiotics*, 8(1), 29–46. https://doi.org/10.1007/s12304-014-9203-2.

Wells, J. C. (2009). Esperanto. In K. Brown & S. Ogilvie (Eds.), *Concise Encyclopedia of Languages of the World* (pp. 375–377). Amsterdam: Elsevier, 1st edition.

Whorf, B. L. (1956). Science and Linguistics. In J. B. Carroll (Ed.), *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* (pp. 207–219). Cambridge, Mass.: MIT Press.

Wikipedia contributors (2019a). Krokodili. *Vikipedio*. https://eo.wikipedia.org/w/index.php?title=Krokodili&oldid=6620887.

Wikipedia contributors (2019b). Reptiliumi. *Vikipedio*. https://eo.wikipedia.org/w/index.php?title=Reptiliumi&oldid=6625438.

Wikipedia contributors (2020a). Kabei. *Vikipedio*. https://eo.wikipedia.org/w/index.php?title=Kabei&oldid=6940800.

Wikipedia contributors (2020b). Kazimierz Bein. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Kazimierz

_Bein&oldid=967113418.

Williams, B. (2005). *Descartes: The Project of Pure Enquiry*. London and New York: Routledge, revised edition.

Wittgenstein, L. (2001). *Tractatus Logico-Philosophicus*. Routledge Classics. London and New York: Routledge.

Wittgenstein, L. (2003a). *Logisch-philosophische Abhandlung =: Tractatus logico-philosophicus*. Number 12 in Edition Suhrkamp. Frankfurt am Main: Suhrkamp.

Wittgenstein, L. (2003b). *Philosophische Untersuchungen*. Number 1372 in Bibliothek Suhrkamp. Frankfurt am Main: Suhrkamp, 1st edition.

Wittgenstein, L. (2009). *Philosophical investigations*. Chichester, West Sussex, U.K. and Malden, Mass.: Wiley-Blackwell, revised 4th edition.

Wrisley, G. (2011). Wittgenstein's Private Language Argument. In M. Bruce & S. Barbone (Eds.), *Just the Arguments* (pp. 350–354). Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444344431.ch94.

Young, J. O. (2018). The Coherence Theory of Truth. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition. https://plato.stanford.edu/archives/fall2018/entries/truth-coherence/.

Zimmer, B. (2009). Skxawng! *The New York Times*. https://www.nytimes.com/2009/12/06/magazine/06FOB-onlanguage-t.html.