The Impact of Underdetermination on the Philosophical Method of Cases

Alexander Seifert

universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

The Impact of Underdetermination on the Philosophical Method of Cases

verfasst von / submitted by

Alexander Wolfgang Seifert von Eichenstark, BSc BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2015 / Vienna 2015

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 944
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	History and Philosophy of Science
Betreut von / Supervisor:	Univ.-Prof. Dr. Martin Kusch

The Impact of Underdetermination on the Philosophical Method of Cases

Introduction

The method of cases is a philosophical tool with a long history, dating back at least to the works of Plato. It allows to test a philosophical theory by conducting an experiment in thought alone. Let's say you have a working theory of the philosophical concept of knowledge: what it means to know something is to have a justified, true belief. Now you test that theory by stipulating a scenario like the story of Smith and Jones from the original Gettier case, which fulfills all the requirements of the traditional theory of knowledge, but where most people *still* want to say that the subject S *does not* know that *p*. These judgments are what we would call *intuitive*, and much philosophical writing assumes that they are not just a trivial consequence of your background theory, but rather hold some evidential weight of their own. As such, our intuitive judgment in the stipulated case serves as counterevidence against the tested theory.

This method has been used for thousands of years, and very heavily so. We find examples in classic antiquity, like Plato's description of Gyges' ring; we find them throughout medieval scholastic philosophy, like Aquinas' use of cannibals to get clear on the concept of resurrection; we find them in early modern philosophy, as with the Cartesian Demon; and of course we find examples abound in contemporary philosophy, such as the Gettier cases, stories of brains

in vats and philosophical zombies. It is easy to see that the method of cases is at the heart of a lot of philosophical theorizing.

For a method so central to the practice of philosophy, one would expect its epistemological grounding to be well understood. Surprisingly, though, for the longest time its reliability was blindly taken for granted. Only recently, beginning in the late 1980s, the faculty of intuition, which is part of the method of cases has come under serious attack. Studies from what now falls under the rubric of 'experimental philosophy' have shown systematic variation in peoples' intuitions, a result that very much threatens the reliability and objectivity of the faculty of intuition and by extension also the method of cases. What followed was a lively debate about the epistemology of intuition, which thrives to this day and deals with questions like: Are intuitions beliefs, or are they mental states sui generis? Are they evidence for deeper philosophical truths, or are they no more than individual psychological states without epistemic import? Is there a wholly rational and – so as not to beg the question – non-intuitive explanation for them having evidential weight and if not, how much of a problem is that? These are some of the issues that are central to the ongoing intuitions-debate.

In my thesis, I want to put the spotlight on another serious shortcoming of thought experiments: the inherent fact that each and every thought experiment can only provide a partial description of a possible world. What I hope to

show is that this underdetermination of every thought experimental scenario opens up the method of cases to various objections. My strategy will be to analyze the debates revolving around two highly influential thought experiments and then demonstrate how each debate can be used to illuminate a different methodological problem. The common thread of these problems is that they are a direct consequence of the thought experiment giving only a partial description of a possible world.

None of these objections are entirely new. However, while some have been wrongfully ignored in the past, others have not been properly recognized for their implications for the method of cases. What my thesis would have to offer, then, is to connect these objections to the array of already existing issues facing the faculty of intuition and the method of cases.

Considering the central role of the method of cases in philosophical practice and the seriousness of the objections, I believe that my thesis offers fresh insights into a truly pressing methodological problem.

❧

The first three chapters of my thesis serve as an introduction for the reader, getting clear on the history, terms, and problems connected with the method of cases. Chapter 1 will show how it has always figured prominently in philosophical discourse. Demonstrating the pervasive use of that method will hopefully prepare the reader to accept the importance of what is at stake here. In

chapter 2 I will provide definitions of the central terms used in this thesis, and I will explain how all of these terms hang together. The crucial insight from this chapter will be to see how the method stands and falls with the faculties of imagination and intuition. Chapter 3 will serve as an overview of the contemporary intuitions-debate to both provide the bigger context of my work and to heighten the reader's sensibility for the range of problems that the method of cases grapples with.

In part I of my thesis I will look at two influential thought experiments and the discussions that revolve around them. Here, I am not so much interested in the philosophical questions posed by these thought experiments; I am rather interested in a formal feature of the debate: where and how intuitions in these cases come apart.

Every debate brings to light a different metaphilosophical problem. By looking at Derek Parfit's Combined Spectrum case in chapter 4, we will see how cognitive limitation poses a problem for a scenario's successful conceivability. The debate around David Chalmers' zombies in chapter 5 will help us tap into the theoretical debate about conceivability and possibility. Chapter 6 then introduces some attempts to save modal intuition in the face of the threats laid out in the previous chapters.

In part II, we will zoom in on Hilary Putnam's Twin Earth scenario (chapter 7), and we will be confronted with the problem of underdetermination at the conceptual level. As an explanation of that problem, chapter 8 introduces us to the theory of meaning finitism, and in chapter 9 I shall lay out the implications of this theory for the method of cases.

Finally, we will draw the conclusions from what I laid out in the previous chapters, and explicate what this means for philosophical practice. By the end of the thesis, the reader should have a sense of the many epistemological difficulties concerning that essential part of the philosopher's toolbox that is the method of cases.

1. Historical Notes on the Method of Cases

The method of cases is a philosophical tool as old as the discipline of philosophy itself. It has figured prominently in philosophical discourse from the past to the present, Plato to Putnam. Everybody who has ever read a Platonic dialogue has encountered the method of cases. In these dialogues, Socrates sets out to answer some profound philosophical problems: What is knowledge? What is justice? What is truth? He confronts laymen and experts with these questions, and when they try to come up with a definition, Socrates' reply frequently takes the form of a counterexample: Suppose that this or that were the case, he would retort, would that be an instance of knowledge/justice/truth?

These counterexamples put the conceptual definition to a test by submitting it to an experiment conducted in thought alone. They are hypothetical cases, which fulfill the definition set up by Plato's interlocutor, but intuitively would be excluded from the concept under examination. This tension is taken to reveal an error in the originally proposed definition.

The method of cases was employed to no lesser extent during the Middle Ages, even though these cases are often not that well-known. St. Augustine uses a thought experiment to demonstrate that one's desire for God is stronger than for every worldly good, be it pleasure, power, or even peace of mind and a good conscience; Thomas Aquinas discusses the case of cannibals resurrected on Judgment Day in order to counter primitive versions of hylemorphism in the question of personal identity; and Duns Scotus cites the hypothetical story of a lamb fearing a "sheep in wolf's clothes" against Avicenna's claim that an intention is literally in the object and linked to sensible forms.

From early modern philosophy, we find a number of thought experiments that even many non-philosophers are familiar with. Galileo refuted the Aristotelean understanding of gravity with an ingenious thought experiment, showing that the Aristotelean conception contains a serious contradiction. Descartes' *genius malignus* illustrated to laymen and philosophers alike the fallibility of (most of our) knowledge. And Laplace's demon (or 'intellect', as he

calls it) is still summoned today to illustrate the philosophical position known as causal determinism.

With the turn of the 19th century, thought experiments became an important tool in the ever more abstract domain of physics. In his influential essay "Über Gedankenexperimente" (Mach, 1897), the Austrian physicist and philosopher Ernst Mach popularized the word 'Gedankenexperiment' (thought experiment), which was coined some 75 years earlier by the Danish physicist and chemist Hans Christian Ørsted (Oersted, 1822). Thought experiments featured prominently in Einstein's arguments for his theories of general and special relativity (Elevator Argument; chasing the beam of light) as well as the interpretation of quantum mechanics (Schrödinger's Cat) and many more.

In the context of the linguistic turn and the following upsurge of analytic philosophy, the method of cases was used heavily to get clear on the definition of our everyday words and philosophical concepts. Edmund Gettier's thought experiment is responsible for what is probably the greatest revolution in epistemology since 2000 years, and the discipline is still struggling with the intuitive insight that Gettier shared some 50 years ago. Extraordinary cases characterize not only the theory of knowledge, but also the philosophical analyses of the human mind, personhood, meaning and reference, ethics, and many more.

❧

Despite some concerns about the use of intuitions, the practice is as widespread as ever, and intuitive judgments feature prominently in many a landmark paper old and new.

In recent years, however, a growing number of empirical studies has added to these theoretical concerns, and together these doubts loom heavily on the use of intuition. A lively debate has broken out, the central points of which will be laid out in the following section.

2. The Debate Around Philosophical Intuitions

Even though philosophers have always relied heavily on the method of cases and the underlying use of intuition, a recent debate has unearthed the many puzzling questions surrounding the faculty of intuition. Questions like: What is the nature of intuitions? How do they transfer evidence? What exactly are they evidence of? Is it the case that we have to rely on foundational intuitions as the basis of our theories? Are intuitions uniform among different groups of people?

The whole debate boils down to one central question: Can intuition enjoy an evidential status (for philosophy) analogous to that enjoyed by perception (for empirical science)? In the words of Ernest Sosa, "intuitions are supposed to play a foundational role in philosophy and other a priori disciplines ... They

are supposed to provide justification relevantly independent of any reasoning, memory, perception, testimony, etc." (Sosa, 2007, p. 45)

Considering that philosophers like George Bealer describe intuitions as grounding the autonomy and authority of philosophy as a discipline, the reader gets a feeling of the imminent danger any serious issue with intuition poses for the whole discipline.

There is not much agreement concerning the nature of philosophical intuition, apart maybe from the very basic assessment that intuitions are "mental states or events in which a proposition seems true in the manner of these propositions." (Pust, 2012) One early account held that intuitions are *kinds of beliefs*. An immediate problem with this view is that sometimes, as in the case of resolved paradoxes, we might have the intuition that a proposition *p* is true even though we do not hold the belief that *p*.

A different theory understands intuitions as *dispositions to believe*. In one of his earlier papers on intuition, Sosa held the view that *S* has the intuition that *p* if and only if *S* is disposed to believe *p* merely on the basis of understanding *p*. This account separates the intuition from the belief, and thus sidesteps the problem of paradoxes. However, still, the definition seems to be too broad: There are very many propositions *p*, which a subject *S* might be disposed to

believe at any one point, but which she is not currently considering. Most philosophers see this element of occurrence as a necessary component of intuition and discard the "disposition to believe" account on that ground.

Honoring that element of occurrence, many philosophers now believe that intuition is a *mental state sui generis*. Sosa calls it an "attraction to assent triggered by considering a proposition consciously with understanding" (Sosa, 2007, p. 60). Bealer further distinguishes philosophical intuitions from physical intuitions when he specifies that S has a "rational intuition that p IFF it intellectually seems to S that necessarily p" (Bealer, 1998, p. 165).

Other philosophers, however, maintain that such phenomenologically mental states *sui generis* do not exist in their own mental life. Williamson for example asserts:

> For myself, I am aware of no intellectual seeming beyond my conscious inclination to believe the Gettier proposition … Of course, dwelling introspectively for long on any belief or inclination to believe has its characteristic phenomenology, but that is the phenomenology of the dwelling, not of what is dwelt upon. These paradigms provide no evidence of intellectual seemings, if the phrase is supposed to mean anything more than intuitions in Lewis's or van Inwagen's sense. (Williamson, 2007, p. 217)

In other words, Williamson thinks that what are usually called 'intuitions' are nothing more than beliefs or inclinations to beliefs that are not distinct in any philosophically interesting way from the more general beliefs or inclinations to believe.

Whatever the precise nature and inner workings of intuition, it is necessary for the method of cases that intuitions are reliable and that they confer evidence. By looking at the source of intuitions, philosophers have tried to see whether or not there could be a reliable indicatorship relation between intuitions and whatever source generates them.

George Bealer links intuitions to determinate concept possession, the idea being that we are all able to (fully) possess the same concepts about knowledge, truth, etc., which allows us to have the same intuitions on these cases. So whenever someone fully possesses the concept of, say, knowledge, then she will have the appropriate Gettier intuition.

One problem with that model is that there is little room for performance errors: diverging intuitions would imply differing concepts. As a reaction to this shortcoming, Sosa defends a competence/virtue-based model of intuition, which is more permissible of performance errors than Bealer's determinate concept possession view. Sosa's model accounts for the fallibility of intuition and can also "allow that paradox-enmeshed proposition contents exert proper attraction, on which one might even base justified intuitive belief. The attraction or belief is justified because it is competent." (Sosa, 2007, p. 59)

Alvin Goldman (Goldman, 2007) dismisses both platonic forms and natural kinds as possible sources because he thinks we have no good reason to believe

that there is such a connection between intuitions and the non-conceptual entities outside the mind. Concepts in the Fregean sense (i. e. abstract entities, graspable by multiple individuals, capable of becoming objects of a faculty of rational intuition) are also problematic because they cannot be unified either by content, phenomenology or by psychological origin.

What gives rise to our intuitions, Goldman thinks, are our personal psychological concepts. That is because there is a constitutive relationship between concepts and the intuitions they generate: "It's part of the nature of concepts … that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept." (ibid., p. 15) For a number of reasons, intuitions are not infallible evidence about that personal concept. However, "a person's application intuitions vis-à-vis their own personal concepts are highly likely to be correct" (ibid., p. 16), so the reliability criterion for evidence-conferring power is met. Whenever intuitions clash, it is because people have slightly different personal psychological concepts. When they overlap, which is the case most of the time, intuitions overlap too.

As with the nature of intuition, also their precise source remains a point of disagreement. Demanding a minimum requirement for reliable sources of knowledge, Robert Cummins points to another problem with giving intuitions evidential status, in that there seems to be no way to calibrate them independently. As Cummins stresses:

> "[A]n observational technique is deemed acceptable just to the extent that it can be relied upon to produce accurate representations or indicators of its targets. This is why observational procedures in general, and instruments in particular, have to be calibrated." (Cummins, 1998, p. 116)

Philosophical intuition could only be calibrated on the assumption that there is some non-intuitive access to its targets:

(P1) One is justified in believing the contents of a putative source of evidence only if one has independent justification for the belief that the putative source is reliable.

(P2) We lack independent justification for the belief that intuitions are reliable.

(C) We are not justified in believing the contents of intuitions.

Independent justification could come only from a relevant theory that is "well enough settled to form the basis of a credible calibration test" (ibid., 118). However, if we already had such a theory, the argument goes, then we wouldn't need the intuitions anymore, leaving intuitions as a source of knowledge either unreliable or useless.

The problem with requiring independent justification of intuition, as Cummins does, is that such a thing is unattainable in principle: if intuition must be calibrated by another source X, then X must be calibrated itself by another source Y, and then the same goes for Y, etc. While skeptics take this to imply that *every* form of justified belief is unattainable in principle, others think that in suitable

circumstances it is nonetheless possible. So we end up with a dilemma: If justified belief of a source without independent calibration is unattainable, then (P1) itself would be unjustified too; and if we don't want to bite the bullet of the skeptic and instead do allow for uncalibrated sources of knowledge, then again we must dispense with (P1).

❧

Putting the calibration objection (most probably representing an unsolvable dilemma) aside, the most urgent objections come from a very unusual and novel source of philosophical reasoning: empirical studies. We find both intrapersonal and interpersonal variation in intuitive judgments, casting doubt on the reliability of the faculty of intuition. There is intrapersonal variation, meaning that one person finds in herself clashing intuitions. A clear case of this phenomenon are paradoxes, but they are rather small in number compared to other intuitive judgments. However, there is also interpersonal variation, meaning that two (or more) people have different intuitions about the same thing. Compared to the interpersonal counterpart, intrapersonal (intrasource) variation is a rather rare beast, so much so, that Pust would even call it a "fact that most of a person's intuitions are not in conflict with one another." (Pust, 2012) As for the latter, however, recent philosophical papers show a growing number of supporters. A new movement called "experimental philosophy" (which even comes with its own manifesto (Knobe & Nichols, 2007), adding to

its urgency) is conducting empirical work to study the "psychological processes underlying people's intuitions about central philosophical issues" (ibid., p. 3).

Experimental philosophers go out in the metaphorical "field" and test thought experiments and philosophical claims with a wide range of subjects to identify the factors that influence a person's intuition. Among the factors, empiricists have found to be co-varying with intuitions are categories such as gender, cultural background, socio-economic status, the order of presentation, etc. – all of which should supposedly be irrelevant in the search for philosophical truths. If these findings are correct, and our intuitive judgments are influenced by such seemingly irrelevant facts, it would seriously undermine the role of intuition in philosophical practice (and, of course, much of the weight of those philosophical positions that are based on key intuitions).

Proponents of intuition have taken different approaches to answering these attacks. One immediate strategy is to identify errors in the studies themselves, for example in the form of problematically small sample sizes or ambiguous questions. Of course, these studies have become more elaborate over the years, and experimental philosophers are working to find new ways to capture accurately a person's intuitions. It seems increasingly clear that at least some of the troubling effects are real and should give defenders of intuition something to think about.

The other strategy in defending intuition is to accept the findings but deny the consequences, i. e. change their interpretation. Such studies don't *really* show clashing intuitions, the rebuttal goes, but rather reveal that different people might operate with slightly different concepts. In response, however, showing that those concepts that we have taken to be more or less universal (like knowledge, meaning, personhood, etc.) are as diverse as studies would then suggest that they are, would be an important result in its own right.

There's a third approach to fending off the implications of said empirical findings by denying that we use intuitions in philosophical theory construction in the first place. It is probably the most forceful rebuttal, as it is the most principled one, sidestepping the problem completely. The idea is this: Even though we *do* use examples and counterexamples to argue for or against philosophical theories, we do not depend on their intuitiveness. Take for example Deutsch's description of the Gettier paper: "Gettier refuted the JTB theory, if he did, … by presenting counterexamples, full stop. Whether these counterexamples are intuitive for anyone is a separate, and purely psychological matter." (Deutsch, 2010, p. 448) Deutsch draws the distinction between *genuine* examples and *intuitive* examples. An example need only be genuine to do its job, it need not be intuitive. The intuition that *p* is not evidence for *p*, but a manifestation of my direct knowledge that *p*. In this way, intuition may be the causal source of a judgment without it being its justificatory source. If Deutsch is right, and we

don’t use intuitions as the justificatory source of our theories, then the problem of unreliable intuitions would be strongly mitigated:

> “in the circumstance in which the philosophical intuitions of two or more cultures clash, how do we decide who is right? … If all we had to rely on in justifying our judgments about philosophical cases were the fact that those judgments are intuitive, cultural variability would put us in a real bind. … The intuitiveness of the judgments is not all we have to rely on, however. We can try to convince those who disagree with us by providing justifications for our intuitions.” (ibid., p. 458)

Similarly, Herman Cappelen thinks that the whole intuitions-debate is mistaken in believing that intuitions are in any way central to the practice of contemporary analytic philosophy. In his book “Philosophy without Intuitions” (Cappelen, 2012) he provides reasons to doubt said assumption. He does so on two accounts: on the one hand by arguing that the whole talk of “intuition” is mostly confused and unhelpful in deciding whether or not intuitions *really* are in play, and on the other hand by giving three hard criteria on the basis of which he judges the occurrence of intuitive judgments in a number of important texts of philosophy, texts that are said to be fundamentally based on intuitions. He finds that in no case intuitive judgments are used to argue for the central point of the respective texts. Cappelen thus concludes that experimental philosophy attacks a practice that doesn’t exist, and reinterpreting their findings as being intuition-agnostic would fail as well because their results would then be trivial, amounting to the claim that judgments in difficult cases are unreliable.

The most valuable part of Cappelen's presentation is probably the second part of his book where he looks at all those philosophical thought experiments. It is the key part in deciding whether or not intuitive judgments actually are central to the practice of contemporary analytic philosophy. He employs three main criteria here:

(F1) *Special phenomenology:* Intuitive judgments are mental states with a distinctive phenomenology.

(F2) *Rock status:* Intuitive judgments have a special epistemic status – they justify, but need not be justified themselves.

(F2.1) *Non-inferential and non-experiental:* Intuitive judgments need no other premises, perception or memory.

(F2.2) *Evidence Recalcitrance:* Even with good arguments against a given intuition, the subject sticks with that intuitive judgment.

(F3) *Based solely on conceptual competence:* Intuitions are a manifestation of competently applying a concept.

The results of Cappelen's book are as bold as they are wide reaching. If he is right, then intuitions are not used in philosophical practice, and the objections from experimental philosophers fall flat. Unfortunately, however, I find a couple of issues with his analysis:

Cappelen hastily excluded (F1) as a useful criterion, because it is hard to detect feelings by looking at texts. "So if no special phenomenology is mentioned,

let's suppose there is none." (Cappelen, 2012, p. 118) This is more of a shortcoming of his methodology, rather than showcasing the absence of intuitive judgments. However, even though many philosophers point to the distinct phenomenology when trying to define what an intuition is, not everyone is on board with this assessment and it is probably not the most essential feature of intuitions, so I regard this issue as being the least urgent one.

In the case of (F2), Cappelen again gravely confines the criterion for methodological reasons, as it is hard to distinguish Rock status from a premise being common ground. "So if there are arguments offered, let's suppose it is not Rock", he explains. (ibid., p. 121) As before, this move has strict methodological reasons, rather than being a real argument for exclusion. Here, however, the theoretical consequence is more problematic, as the special epistemic status is undoubtedly the most important characteristic the faculty of intuition holds for philosophical discourse.

Let's look at one of, if not *the* most famous philosophical intuitions, the so-called Gettier intuition. (Gettier, 1963) cites three definitions (listed as a, b, c) for knowledge and goes on to give two examples that refute these definitions.

Gettier does indeed provide a short remark that resembles an explanation:

"[I]t is equally clear that Smith does not *know* that [key proposition] (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in

(e) on a count of the coins in Jone's pocket, whom he falsely believes to be the man who will get the job." (Gettier, 1963, p. 122).

What he says is that the proposition is true by accident, which is the foundation for later "justified true belief + x" theories of knowledge. However, this remark is extremely brief and parenthetical almost, so much so that I don't think Gettier intends it to be a proper philosophical argument. It becomes even more apparent that it is cases (with the accompanying intuition), and not the explanation, which do the argumentative work when we look at the concluding paragraph of his paper:

> "These two examples show that (a) does not state a *sufficient* condition for someone's knowing a given proposition. The same cases, with appropriate changes, will suffice to show that neither definition (b) nor definition (c) do so either." (Gettier, 1963, p. 123)

The relative space that the cases occupy as well as Gettier's wording clearly point towards the intuition rather than the explanation as the key driver of his argument.

As far as (F3) goes, Cappelen argues that this can't be what philosophers are doing because the enterprise of conceptual analysis as a whole is ridden with problems. While I do agree with his critical assessment of that philosophical practice, I don't agree with his handling of (F1) and (F2) and thus disagree also with the results of his close reading, namely the view that philosophers do not engage in conceptual analysis by using intuition. Being unconvinced of his

main argument, I *do* think that this is exactly what happens, even if the practice is riddled with problems, pointing to consequential limitations of the method of cases. In the later part of my thesis, I devote a whole chapter to that problem. However, before we get there, I will look at the role intuition plays in philosophical thought experiments, and show how it is not only the conceptual intuition that is problematic, but that every thought experiment implicitly relies on a modal intuition as well, which suffers from its very own limitations.

3. The Role of Intuition in Philosophy

The method of cases can be represented as a two-step process: You first conceive of a certain scenario, and then in the context of that scenario test your intuition in light of a certain question.

Both steps correspond to a specific kind of intuition: a *modal intuition* is used to stipulate the possibility of the world under consideration, and a *conceptual intuition* is used to answer the philosophical question that the case is supposed to shed light on. Both the modal intuition as well as the conceptual intuition face certain problems that can be subsumed under the broad rubric of underdetermination:

In the first instance, our imagination provides us with some (supposedly) possible world, but only as long as no contradiction is detected. The problem is that in one way the faculty of imagination is almost limitless, and in another

way it is very limited. Our imagination seems almost limitless in terms of subject matter, while at the same time due to our limited human nature, we imagine these possible worlds only in very little detail. This underdetermination of possible worlds bears the result that many worlds that seem conceivable often turn out to involve a contradiction and thus are metaphysically impossible. The ancient Greeks have found it conceivable that stars are holes in the sky; some people find it conceivable that Hesperus is not Phosphorus; and a number of contemporary Mathematicians find it conceivable that the Goldbach conjecture is wrong. In each case, the content that is being conceived could very well be metaphysically impossible; nevertheless, many people have no problem conceiving of it.

Like modal intuitions, conceptual intuitions too face the problem of underdetermination, because the concepts they are dealing with are themselves underdetermined and subtly change through being used in a language community. Modern psychological research shows that our concepts are not represented by necessary and sufficient conditions but rather follow a prototype model of word storage, with the set of prototypes being mutable. Drawing the consequences of such a model for conceptual intuition promises to explain the instability of intuition that has been the subject of a lively debate since the late 1980's.

In the following two parts, I will discuss these two kinds of intuitions and the limitations each of them faces due to their being underdetermined. Part I deals with the first step in the method of cases, namely the modal intuition, while Part II will go into detail regarding the second step, the conceptual intuition.

Starting with Chapter 4 I will ease into the problem of conceiving of a certain thought experimental scenario by providing a close reading of one thought experiment from Parfit's *Reasons and Persons*. After familiarizing the reader with the problem, I will follow up in Chapter 5 with a theoretical discussion. I will first say more about the role conceivability plays in the method of cases before I discuss one prominent attempt to solidify the link between conceivability and possibility. I will argue why I think this attempt fails and why philosophers are left with a rather unreliable notion of conceivability. In reaction to this failed attempt, I will use the final chapter of this part to briefly discuss some possible ways to cordon off good and bad cases of conceivability.

I will start the second part with an examination of the debate around Putnam's Twin Earth scenario (Chapter 7), with a special focus on conceptual underdetermination. I will again follow up with a theoretical discussion in Chapter 8, where I introduce the reader to the theory of meaning finitism from the sociology of knowledge and show what that theory holds for the method of cases. I will conclude the second part by connecting the meaning finitist view to recent work by epistemologists and philosophers of intuition. By the end of this

part, we have re-interpreted the method of cases through the lens of the sociology of scientific knowledge and gained a better understanding of why we should expect intuitions to clash and what that means for conceptual analysis.

Part I: Modal intuition and Cognitive Limitation

Philosophers often make use of thought experiments involving situations that could never come up in the real world. These cases feature swamp monsters, zombies, teleportation, and many other figments of imagination. Stories like these are stipulated in order to get clear on the nature of various concepts like knowledge, consciousness, or personal identity. Such arguments presuppose that whatever you can conceive of is (metaphysically) possible. This intuitively plausible link between conceivability and possibility is captured in the traditional view that *conceivability entails possibility*, which owes itself to the phenomenological observation that "conceiving involves the appearance of possibility", as Stephen Yablo put it (Yablo, 1993, p. 5), or "conceiving that p involves intuiting that p is possible", to put the same point in the words of George Bealer. (Bealer, 2004, p. 15) This view goes back at least to David Hume, who famously claimed that it is

> "an established maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence*, or in other words, *that nothing we imagine is absolutely impossible*." – (Hume, 2000, I. ii. 2, via Gendler & Hawthorne (2002, p. 17), emphasis in the original)

So when a philosopher says that something is *conceivable*, what he's *really* trying to establish is that it is *possible*. Accordingly, we can now further explicate the role of conceivability in the method of cases by expanding the two-step picture of given before into a three-partite structure (Chalmers, 2002, p. 145):

1. first, we have an *epistemic claim* (about what can be known or conceived)
2. from there we proceed to a *modal claim* (about what is possible or necessary),
3. and then, finally, to a *metaphysical claim* (about the nature of things in the world).

One classic example from the history of philosophy comes from Descartes, who argued for his mind–body dualism on the basis of no more than mere conceivability. The argument mirrors the three steps outlined above:

1. it is conceivable that mind exists without body, and conceivability implies possibility,
2. so it is possible that mind exists without body,
3. thus, it is actual that mind and body are distinct.

The argument appears to be valid. (2) follows immediately from (1), while (3) follows from (2) according to a more general, very powerful principle: If you can imagine *a*'s existence without *b*'s existence, then it is possible that *a* exists without *b*. And as nothing can exist without itself, it follows that *a* and *b* must be distinct (paraphrase taken from Gendler & Hawthorne, 2002, 11f).

But what about assumption (1)? It consists of two parts, both equally problematic: Is it *really* conceivable, that mind exist without body? And *does* conceivability imply possibility? In the following chapter, we will look at a famous thought experiment by Derek Parfit, which at first glance seems to be perfectly

conceivable, but for which I argue that it is quite probably incoherent and most certainly not as clear as it seems, casting doubt on its possibility.

4. Parfit's Combined Spectrum Case

In this chapter, we look at one of Derek Parfit's thought experiments from his seminal book *Reasons and Persons* (Parfit, 1986), which deals with the concepts of personal identity and psychological continuity. The book is full of strange, science-fiction style cases, making it a perfect source to study the limits of conceivability.

In *Reasons and Persons* Parfit introduces three separate but closely interrelated thought experiments, that together help to make one larger point. These are the cases of the so-called *psychological spectrum*, the *physical spectrum*, and a mixture of these two, the *combined spectrum*. The most interesting one for us is the combined spectrum case, which goes like this:

Suppose I am taken hostage by a mad scientist. This scientist has a teleportation device that first records my blueprint, then destroys me, and then builds an exact replica of me from that blueprint. The machine has a control panel with a large number of switches. Proportional to the number of switches that he flips before pressing the button, the resulting Replica will be some combination of Greta Garbo and me. Now consider the following two extreme ends

of a spectrum of cases, each of which can be seen as a separate thought experiment (Parfit, 1986, p. 236f):

> "At the near end of this spectrum is the normal case in which a future person would be fully continuous with me as I am now, both physically and psychologically. This person would be me in just the way that, in my actual life, it will be me who wakes up tomorrow. At the far end of this spectrum the resulting person would have no continuity with me as I am now, either physically or psychologically. In this case the scientist would destroy my brain and body, and then create, out of new organic matter, a perfect replica of [...] Greta Garbo."

Beside those extreme cases, every combination of Garbo and me will be possible, depending on how many switches the scientist flips.

Parfit uses this thought experiment – or range of thought experiments, rather – to show that competing, non-reductionist views of personal identity hang on implausible suppositions. The main such supposition is that our identity must be determinate, or in other words: there is always an answer to the question "Is some future person P identical to me?" However, Greta Garbo is clearly not me, and if she is not, then somewhere along the spectrum, there must be a sharp borderline. Given that each case along the spectrum differs only so slightly, however, it is implausible that such a small change should have such a big impact, and thus the supposition of determinate identity is probably wrong.

So much for Parfit's use and interpretation of his Combined Spectrum case. In the next section, we will look at the historical reception of Parfit's thought experiments before I present my own reasons to doubt the possibility of the case at hand.

4.1. Reactions to Parfit

Looking at book reviews and discussions of *Reasons and Persons* shows the unease philosophers felt with Parfit's examples. However, while his cases certainly seemed to be noteworthy, hardly anyone took a more principled, critical stance toward the type of thought experiments that Parfit employs. Many authors mention the fact that Parfit's examples are highly imaginative (Adams, 1989; Shoemaker, 1985; Simpson, 1985; Wolf, 1986). Moreover, some seemed to be suspicious of these "science fiction style" experiments. However, nobody pointed to a particular problem with Parfit's examples. Robert Adams, for instance, praises the "wealth of ingenious and fascinating examples" (Adams, 1989, p. 439), describing them as being "of a predominantly science fiction character" (Adams, 1989, p. 454). His one critical remark spans no more than two sentences:

> "[Parfit] begins with intuitions that I think draw their power (though he might deny it) from the difference between these examples and normal, clear cases of personal identity, and then tries to use these intuitions to talk us into assimilating the normal to the abnormal case. But we cannot pull ourselves up by our own intuitive bootstraps." – (Adams, 1989, p. 466)

Peter Simpson offers a similar verdict:

> "I am not persuaded overall because [Parfit's] argument relies on two beliefs I am not inclined to accept and which Parfit has given me no reason to accept. These are that there is no alternative between Cartesian egos and reductionism, and *that the replacement experiments he imagines are in principle, if not in practice, possible*." – (Simpson (1985, p. 372), emphasis added)

In a move that is symptomatic of the reactions to Parfit's cases, Simpson hedges this previous claim, concluding his review with the remark that

> "despite these reservations the arguments remain powerful and stimulating. … [The book's] conclusions may be disturbing but there is no denying the power of their presentation." (ibid.)

None of these authors go into detail regarding their issues with Parfit's cases, and nobody provides an argument for why "we cannot pull ourselves up by our own intuitive bootstraps", to borrow Adams' phrase. Without any such kind of argument, we must interpret those authors as taking Parfit's cases to be strange but unproblematic, or at most diagnosing a clash of intuitions about the possibility of these science fiction cases.

We do not have to suspend judgment and point to a clash of intuitions, however. In the following section, I will try to make the case that at least some of Parfit's cases are impossible, even if they seem conceivable at first glance.

4.2. Incoherencies of the Stipulated Case

I will argue that in this particular group of thought experiments we have reasons to doubt the conceivability of most of the cases under consideration. As we are dealing with intuitions, I will begin with a short introspective remark.

Imagining cases from those two extreme ends is rather easy: I just imagine either myself or Greta Garbo. When it comes to the intermediate situations, however, I find it so hard to imagine what those cases would actually be like, that my intuition starts to give in.

There are two things to note here: First, remember that the argument depends crucially on there being a fine-grained (discrete) continuum of cases, I will call this the *fine-grained continuum assumption*. Parfit needs this assumption in order to appeal to the intuition that the difference between two neighboring cases is so small that it could not effect a change from me being identical to the Replica in the one case to the neighboring case where I cease to exist, and Greta Garbo comes into existence.

Second, his argument depends on the assumption of surgical replaceability of those tiny parts of both the physical and the psychological. Let's call this the *atomic replaceability assumption*.

With this second assumption on the table, Parfit's case loses its intuitive appeal for me, as I can no longer assert that the atomic replaceability assumption

– the idea of surgically replacing single cognitions – is truly coherent. Can I imagine a person that is just like me except that it has Greta Garbo's intentions X, Y, and Z, without also having some of her other memories, desires or intentions? How could I hold, for example, Greta Garbo's memory of a dinner with her friends, the way Greta Garbo remembers it, without also thinking that those people are *my* friends, without remembering how I felt in that situation and without making any associations with other memories, desires or intentions?

One might try to respond by showing how what I said is perfectly compatible with Parfit's thought experiment. Such a response could go like this: "Well if intentions, desires and memory hang together in clusters then we'll simply replace cluster by cluster." – This too is problematic for two reasons:

First, it is not clear that replacing whole clusters is coherently possible either because the clusters need not overlap so neatly. By replacing one me-cluster for one Garbo-cluster, I might, for example, import some new intentions that contradict other previously held intentions or desires. While it is of course to some degree possible to hold contradictory views/intentions/desires, it is implausible to suppose that a person with an infinitely high degree of internal contradiction can be coherently conceived.

Reason number two: Even if the replacement of clusters *were* possible, Parfit's argument loses its plausibility, because by loosening the atomic replaceability

assumption, his fine-grained continuum assumption loses plausibility. This is because the replaceability assumption derives much of its plausibility from the fine-grainedness of the spectrum. As soon as we start to replace larger clusters instead of atomic cognitions, the fine-grained continuum crumbles, and the competing intuition becomes plausible again that maybe there *is* a sharp borderline where I "stop being me" after all. The mad scientist might have replaced one large chunk of my mental life with one large chunk of Garbo's mental life, and that might have tipped the scale.

Another possible counterstrategy might be to retell the thought experiment as a story that connects each of the points on the continuum by rearranging them into separate time-slices of one single person. In this new narrative, my current self would undergo a slow but continuous transformation into Greta Garbo, with each step being conceivable on its own.

The trick of this re-description is that it takes the very real and actual phenomenon of gradual personal development over time and uses it to make plausible the possibility of the gradual transformation in the original Combined Spectrum case. I would object that there is a critical difference in these two descriptions: The Combined Spectrum case assumes two very different persons at each end of the spectrum because it gains its force from the intuition that there is no sharp borderline where I "stop being me". The re-description drops this

key feature, making the whole scenario possible at the price of losing its argumentative force. Without this key feature, the analogy doesn't work, and the possibility of the original case remains questionable.

A third possible reaction to my argument might point to the fact that a larger number of cognitions are so isolated that we could replace them easily: factual beliefs like *Cicero lived from 106-43 BC*, emotions like *I currently feel happy* or intentions like *I intend to buy a car* might come to mind.

I have two responses: First of all, I am not sure that even trivial cognitions like these are so easily interchangeable: my factual beliefs about Cicero's birth and death are connected with my beliefs that he was a contemporary of Caesar, that he lived during the Roman civil wars, and so forth. Equally, an intention like *I intend to buy a car* cannot so easily be added to my belief that *I am in debt*, my intention to save money or my conviction to live a frugal life and have a low carbon footprint.

Second, not every cognition is equally integral to my personal identity. Some cognitions are more central to me being me than others, and I would argue that they also cluster together more strongly. They revolve around friends and family, life goals, strong convictions and similar things that we attach much importance to and that we integrate into our conception of our life and our personality. For this reason, those cognitions are much more strongly intercon-

nected, much more clustered together. So if this is true, it would again undermine the fine-grained continuum assumption and thus have the effect that replacing such a cluster effects a change in someone's personality, undermining Parfit's thought experiment.

Summarizing my analysis of Parfit's case: If I am not alone in my trouble imagining even superficially how most of these cases would turn out, this would throw doubt on Parfit's assumption that you could coherently conceive something like the Combined Spectrum. Of course if the described scenarios are not even conceivable, you cannot build an argument on their possibility, and so Parfit's argument is blocked before it can even get off the ground.

❧

In the next chapter, we will dive into the technical details of conceivability and possibility, together with David Chalmer's theory of modal rationalism, which links these two concepts together in an attempt to lay the epistemological foundation for philosophical thought experiments.

5. From Conceivability to Possibility

As I laid out in the previous chapter, the debate on Parfit offers little in regards to the questionable epistemic value of his cases. I then presented reasons to doubt the possibility of Parfit's Combined Spectrum case. We now turn to a theoretical discussion of the basic link between conceivability and possibility.

Due to the likeness of the problem, the zombie debate is an excellent resource for thinking about Parfit's thought experiments, as it illustrates the problematic relationship between actuality, conceivability and possibility.

You could sum up the positions from the debate in the following way: Hardly anyone thinks that zombies actually exist, but many philosophers hold that they are conceivable, and some believe that they are possible. The reactions to Parfit are very similar: No philosopher thinks that Parfit's cases about teleportation, division, or a combined spectrum actually exist, but many of them hold that they are conceivable, and some believe that they are possible.

In this chapter, we will see how conceivability and possibility are connected, and also how – as I will argue – they can come apart as in the cases of Parfit's Combined Spectrum and Chalmers' zombies.

5.1. Does Conceivability Entail Possibility?

> "Alice laughed. 'There's no use trying,' she said; 'one can't believe impossible things.'
>
> 'I dare say you haven't had much practice,' said the Queen. 'When I was your age, I always did it for half an hour a day. Why, sometimes, I've believed as many as six impossible things before breakfast.'"
>
> – Lewis Carroll, *Through the Looking Glass*

In the last two decades the link between conceivability and possibility, which for a long time was taken for granted, has been called into question: maybe we are able to (more or less easily) conceive of situations that do not correspond

to a possible world, i. e. are impossible. This worry is one of the central points in the zombie debate from the philosophy of mind. In philosophy, zombies are defined as beings that are physically exactly like us, only they lack conscious experience; or in Robert Stalnaker's almost poetic phrasing: "The sun shines in such [zombie] worlds, but the lights are out in the minds of the unfortunate creatures who live in them." (Stalnaker, 2002, p. 385) David Chalmers' so-called zombie argument is directed against the hypothesis that consciousness logically supervenes on the physical world, which is a consequence of materialism, i. e. the claim that

> "Among worlds where no natural properties alien to our world are instantiated, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates." – (Lewis, 1983, p. 364)

The argument's basic idea is this: Considering that we can imagine a zombie world, it should be at least metaphysically possible for such a world to exist, which means that physical facts alone cannot be taken to account for all the phenomena and materialism is false. In his book *The Conscious Mind*, David Chalmers makes heavy use of the zombie case to argue for his version of Dualism. His zombie argument against materialism can be laid out as follows:

(P0) In our world, there are conscious experiences. (From introspection we know that we are not zombies.)

(P1) We can conceive of a logically possible zombie world, i. e. a world physically identical to ours, in which the positive facts about consciousness in our world do not hold.

(P2) Such zombie world is at least metaphysically possible.

(C) Therefore, facts about consciousness are *further facts* about our world, *over and above the physical* facts and materialism is false.

We can see how the argument follows the same three-partite structure as the Cartesian argument for mind-body dualism outlined above (only with the assumption (P0) added for clarification). Here again, the argument seems to be valid. Here again, the epistemic claim about conceivability (P1) is doing the whole work of the argument. Moreover, it is this epistemic claim that became subject to much controversy in the philosophy of mind.

For Chalmers, the conceivability of a zombie world is simply an intuitive truth. "In some ways", he says, "an assertion of this logical possibility comes down to a brute intuition" (Chalmers, 1996, p. 96). If authors like Colin McGinn or Daniel Dennett are right, however, then the truth of his premise is highly questionable: They argue that in complex scenarios it may *seem* to a subject that S is conceivable, but that's only because the human mind is too limited to ponder all the relevant details at the same time, and is prone to overlook subtle incoherencies. (McGinn, 1989; Dennett, 1984) Chalmers' program of *modal rationalism* sets out to give a comprehensive rebuttal of these doubts, so let us take a closer look at that program.

5.2. Modal Rationalism and Ideal Positive Conceivability

Trying to establish the link between conceivability and possibility, David Chalmers worked out a taxonomy of different kinds of conceivability. (Chalmers, 2002) He carves out three dichotomies:

- positive versus negative conceivability
- prima facie versus ideal conceivability
- primary/epistemic versus secondary/subjunctive conceivability

Here's how Chalmers defines those concepts:

S is *negatively conceivable* for a subject when that subject cannot rule out S on a priori grounds. It is *positively conceivable* when one can modally imagine a situation that one takes to be coherent, and that one takes to verify S.

S is *prima facie conceivable* for a subject when that subject cannot (after consideration) detect any contradiction in the hypothesis expressed by S. It is *ideally conceivable* when S is conceivable on ideal rational reflection.

S is *primarily conceivable* when it is conceivable that S is actually the case. S is *secondarily conceivable* when S conceivably might have been the case.

As these features are independent of each other, their combination gives rise to 2×2×2 = 8 different types of judgments (prima face secondary negative conceivability, etc.). Looking at the taxonomy, a few points become immediately clear:

1. *Prima facie conceivability is an imperfect guide to possibility.* This is because secunda facie conceivability and ideal conceivability represent subsequently higher degrees of undefeatability.

2. *Positive conceivability is a better guide to possibility than negative conceivability.* This is because positive conceivability implies negative conceivability but not the other way around.

3. *Primary conceivability is an imperfect guide to secondary possibility.* This is because primary and secondary possibility are independent of each other.

This leaves the following two central conceivability–possibility theses, which are of principal interest to the kind of modal rationalism Chalmers wants to defend:

1. Ideal primary positive conceivability entails primary possibility.

2. Ideal primary negative conceivability entails primary possibility.

The qualifications of (a) ideal and (b) positive conceivability are crucial to linking conceivability and possibility. A conceiving must be ideal because prima facie conceivability is epistemically too weak to entail possibility (see 1. above). Furthermore, a conceiving must also (at least indirectly) be positive in kind, because ideal negative conceivability entails primary possibility only transitively: directly it only entails ideal positive conceivability (a claim that needs a pretty complicated and extensive argument given in Chalmers (2002)), and via positive conceivability it entails primary possibility – thus its power is parasitical upon the notion of positive conceivability.

According to Chalmers, ideal positive conceivings are what we arrive at in successful thought experiments. He describes the typical philosophical thought experiment as starting with prima facie positive conceivability: A subject broadly imagines a certain scenario, leaving unspecified fine details like microphysical facts about the world. Only the most essential features are specified, which the subject then judges to verify S, concluding that the remaining details are not crucial to the experiment. Not being crucial means that they can be filled in to yield a "full, coherent conception of a situation that verifies S" (Chalmers, 2002, p. 153f). This prima facie judgment about which facts are relevant and which are irrelevant for the intended purpose must be correct so that S is ideally positively conceivable: "If better reasoning would reveal that the details cannot be coherently filled in, or that the situation does not truly verify S, then the thought experiment will typically fail in its purpose." (ibid.) Provided that the prima facie judgment is not defeatable in this way, S is ideally positively conceivable, and the thought experiment succeeds.

Shortly after, he notes the precedence of positive over negative conceivability in our actual philosophical discourse:

> "Positive conceivability, rather than negative conceivability, seems to be what most philosophers have had in mind when discussing conceivability. It is positive conceivability … that reflects the practice in the method of conceivability as used in contemporary philosophical thought experiments." (Chalmers, 2002, p. 155)

In what follows I will provide a critique of Chalmers' concepts of positive and ideal conceivability, bearing the result that we can only count on prima facie negative conceivability in actual philosophical practice. If my argument is correct, then conceivability claims in the method of cases are too weak to establish the link to (primary) metaphysical possibility, rendering the whole method unreliable.

5.2.1. Positive Conceivability is Unidentifiable

While the concept of negative conceivability is rather clearly and uncontroversially defined, Chalmers' notion of positive conceivability lacks this clarity of exposition. He himself admits that it "cannot be considered a reductive definition. At best, it is something of a clarification." Nevertheless, he adds, "there seems to be a reasonably clear intuitive notion in the vicinity, of which most people seem to have a grasp." (Chalmers, 2002, p. 156)

Chalmers uses a familiar Cartesian characterization when he describes positive conceivability as the type of conceivability "that corresponds to the sort of clear and distinct modal intuition invoked by Descartes, and that reflects the practice in the method of conceivability as used in contemporary philosophical thought experiments." (Chalmers, 2002, p. 155)

Could it be *clear and distinct* modal intuition, which distinguishes positive from negative conceivability, and which is able to provide the link to metaphysical possibility? I see two difficulties with this: For this kind of intuition to truly be

useful for the method of conceivability as used in contemporary philosophical thought experiments, it must be both identifiable and reliable. Accordingly, Chalmers needs to provide answers to the following two question:

1. How can we identify clear and distinct modal intuitions as such?
2. Why would clear and distinct modal intuitions lead to possibility? What establishes that link?

Both of these problems are not specific to Chalmers' account but can already be found in the discussion of the Cartesian model of clear and distinct intuition, as we will see now.

The first of the two problems was raised by Descartes' contemporary Pierre Gassendi, who complained that

> "the difficulty does not seem to be about whether we must clearly and distinctly understand something if we are to avoid error, but about what possible skill or method will permit us to discover that our understanding is so clear and distinct as to be true and to make it impossible that we are mistaken." – (Descartes, Cottingham, Stoothoff, & Murdoch, 1984, p. 221)

Gendler and Hawthorne describe Descartes' reaction to this line of attack as "impatient", and indeed his reply is short and dismissive: If only the objector engaged in introspection, he would notice his obvious capacity to identify instances in which his understanding or perception is transparently clear. In these cases "perceptions are so transparently clear and at the same time so simple that we cannot even think of them without believing them to be true." (ibid., p. 104)

Chalmers makes a very similar move: Characterizing positive imaginability, he says that although he can give no precise definition, there is "a reasonably clear intuitive notion in the vicinity, which most people seem to have a grasp on" (Chalmers, 2002, p. 156), pointing to the concept of clearly and distinctly modally imagining a situation. I am not entirely sure how we should understand "clear and distinct intuition", but it probably corresponds to what I understand as the strength of an intuition. Engaging in introspection, I notice that there are indeed cases, where I have intuitions that seem to be very strong and in a way *clear* to me. Again, the prototypical candidate here is the original Gettier case. In other instances, the strength of my intuition is rather weak and its direction easily swayed by theoretical consideration: suddenly I find myself with a very different intuition, in direct opposition to my initial reaction. However, these are conceptual intuitions and are not related to the possibility of worlds, they simply reflect my feeling towards the application of a concept. When it comes to modal intuitions, though, I am not sure how helpful a criterion this is. If we take the zombie case, for example, we find clashing intuitions with the reactions in the literature ranging from "possible" to "impossible", rendering positive conceivability ostensively unreliable.

In terms of clarification, Chalmers circumscribes positive conceivability with the notion of imagination. He suggests to put the varieties of positive conceivability under the broad rubric of *imagination*: "to positively conceive of a situ-

ation is to imagine (in some sense) a specific configuration of objects and properties." (ibid., p. 150) He goes on to list two different notions of imagination: *perceptual imagination* and *modal imagination*. Perceptual imagination requires that the subject "has a perceptual mental image that represents S as being the case." He gives the example of imagining that a pig flies by forming a visual picture of a flying pig, meaning an image that relevantly resembles a visual experience as of a flying pig. If we were to do that, we could say that the imagined situation verifies 'Pigs fly'.

So far so good. However, this is not Chalmers' core notion of imagination and conceivability. He is more interested in modal imagination, which is taken to mean that

> "one has a positive intuition of a certain configuration within a world, and takes that configuration to satisfy a certain description. … We can say that an imagined situation verifies S when reflection on the situation reveals it as a situation in which S. Understood this way, verification is a broadly epistemic relation, tied to certain rational processes. Importantly, verification is stronger than a mere evidential relation." – (ibid., p. 151)

Bringing the epistemic notion of 'verification' into play, Chalmers' imports another epistemic concept: modal imagination, if it shall lead to metaphysical possibility, is now intimately linked to coherence, as can be seen from the following key passage on positive conceivability:

> "To avoid cases [where one can modally imagine S when S invokes an a priori contradiction], one can isolate a notion of coherent modal imagination, and

> hold that S is positively conceivable when one can coherently modally imagine a situation that verifies S. A situation is coherently imagined when it is possible to fill in arbitrary details in the imagined situation such that no contradiction reveals itself. To coherently imagine a situation that verifies S, one must be able to coherently imagine a situation such that reasoning about the imagined situation reveals it as a situation that verifies S." – (ibid., p. 153)

Supplementing imagination with the concept of coherence, Chalmers arrives at this central definition: "S is positively conceivable when it is coherently modally imaginable." (ibid.)

Positively conceiving of S now means the same thing as coherently imagining S, or in other words that "it is possible to fill in arbitrary details … such that no contradiction reveals itself." (ibid.) We end up with a definition of positive conceivability that is dangerously similar to the definition of negative conceivability, which held that "S is negatively conceivable when S is not ruled out a priori, or when there is no (apparent) contradiction in S." (ibid., p. 149) After all, if there is no contradiction in S, then it is possible to fill in arbitrary details; and conversely, if you can coherently fill in arbitrary details, then it simply means that there is no contradiction in S. So in the end the distinction between positive and negative conceivability collapses and we are left only with the more clearly defined notion of negative conceivability.

5.2.2. Ideal Conceivability is Impractical

At this point in the dialectic, we are left with at most *negative* ideal conceivability. If it were possible to save as much, the method of cases would be well

off: The second of the two central theses of modal rationalism holds that ideal primary negative conceivability entails primary possibility. This is true because ideal *negative* conceivability entails ideal *positive* conceivability and thus, transitively, also entails primary possibility.

Let us for a brief moment again grant Chalmers his distinction between positive and negative conceivability. For ideal judgments, the entailment above would then be true. Every kind of positive conceivability – both ideal and prima facie – entails the respective version of negative conceivability. However, as Chalmers himself points out, for prima facie judgments, the reverse is not always the case. Many statements are prima facie negatively conceivable without being prima facie positively conceivable. For example, many complex mathematical statements are such that one cannot rule out their truth, but one also cannot imagine any situation (any part of a world) that would verify them. The same holds for statements in other a priori domains. Interestingly, Chalmers notes that "Even in empirical domains, it may be that one cannot rule out M, but one cannot conceive of a situation in which M, due to limited powers of imagination, for example." (ibid., p. 155)

With the entailment running only in the direction from prima facie positive to prima facie negative conceivability, Chalmers would need at least prima facie positive conceivability if he wants to save modal rationalism. However, with the distinction between positive and negative conceivability dissolved in our

previous section and considering that – as Chalmers himself admits – prima facie conceivability is a weak guide to possibility, he needs ideal negative conceivability for his argument to succeed.

Again, a version of this problem was first brought up by a contemporary of Descartes. In the *Second Set of Objections*, Marin Mersenne asks: "how can you establish with certainty that you are not deceived, or capable of being deceived, in matters which you think you know clearly and distinctly?" (Descartes, Cottingham, Stoothoff, & Murdoch, 1984, p. 90) The same question haunts us when reading Chalmers' argumentation: He provides no reason to believe that modal intuition, even if clear and distinct, would be a good guide to metaphysical possibility. As Chalmers nowhere explicitly deals with this issue, here again, we will take a look at how Descartes tried to address that problem.

In the *Second Set of Replies* Descartes writes: "Self-contradictoriness in our concepts arises merely from their obscurity and confusion." (ibid., p. 108) If this is the case, then concepts which are not obscure and confused are not self-contradictory. Replacing the negations with their positive antonyms, we take Descartes as saying that clear and distinct concepts are concepts of something that's possible.

I find this assertion to be highly questionable: Why should self-contradictoriness arise *merely* from obscure and confused concepts? Why can't there be errors in clear and distinct concepts? At this point in Descartes' argument, we

encounter the proverbial deus ex machina, the benevolent God, who would not allow errors arising form clear and distinct concepts: It is contradictory to suppose that "anything should be created by him which positively tends towards falsehood" (ibid., p. 103). The problem with this line of argumentation is of course that for anyone not accepting the dogma of divine benevolence, his explanation would not be too convincing.

So for practical purposes, we have one big problem with ideal conceivability (understood as the type of conceiving where filling in the details would never reveal any kind of incoherence): You cannot distinguish cases of mere prima facie conceivability from instances of ideal conceivability, or in other words: you can never know if you have imagined a *truly* coherent situation that *truly* verifies S. Only in retrospect, when your conception has been falsified, can you tell that the situation *wasn't* ideally conceived. This is an epistemic consequence of our finite human condition. Chalmers recognizes this problem:

> "if we are looking for a notion of conceivability such that conceivability tracks possibility perfectly, we must focus on ideal conceivability. In this sense conceivability is not a merely psychological notion; it is a *rational* notion, in much the same way that a priority and rational entailment are rational notions. **If there is to be a plausible epistemic/modal bridge, it will be a bridge between the rational and modal domains.**" (Chalmers (2002, p. 160), bold emphasis added)

Chalmer's strategy of stablishing a link between conceivability and possibility through the addition of an ideally rational component is problematic: because

his definition of positive conceivability includes the faculty of imagination (which is a psychological notion) appealing to *ideal* rationality should be off limits here. Instead, he appeals to ideal rationality, sidelining the problem of human fallibility. By divorcing conceivability from its psychological aspects, Chalmers trivializes the link between conceivability and possibility, thus leaving the original problem unanswered. This can be seen by looking more closely at the kind of possibility we are after: Metaphysical possibility covers the range of ways the world might have been, how "God might have made things". It is more narrow than logical possibility in that it not only excludes what is logically impossible but also what is conceptually impossible. An example would be things that are both red and non-extended: according to our concepts, if something has a color it needs to have an extension, so it is conceptually/metaphysically impossible.

If conceiving and possibility both were understood as substantially rational notions, Chalmers would run the risk of trivializing the link between the two concepts. By making his main argument only for the ideal notions of conceivability he crosses that line and voids the relevance of his argument. Recall Chalmers' definition of ideal conceivability:

> "S is ideally conceivable when there is a possible subject for whom S is prima facie conceivable, with justification that is undefeatable by further reasoning." (ibid., p. 148)

For such strong cases, conceiving that S trivially becomes conceiving *as possible*. This, of course, is no acceptable move for the skeptic of the Zombie dialectic, who is more concerned about the shortcomings of conceivability as a psychological notion, than in the hardened, impeccable version of conceivability that Chalmers constructs. Moreover, this worry is very much warranted, I think, because after all, Chalmers' notion would not reflect "the practice in the method of conceivability as used in contemporary philosophical thought experiments." (ibid., p. 155)

❧

Let's recap. Chalmers sets out to answer very practical doubts about the method of conceivability: How can we, as philosophers, be confident that when we do a thought experiment and conceive of a situation, this situation is indeed possible? But in trying to come up with an answer, what he effectively does, is that he rips away any kind of psychogenetic flaws in the notion of conceivability until only an ideal observer is worthy of conceiving in such a way, and thus Chalmers arrives at a notion that (without his realization) trivially implies possibility and is irrelevant for actual philosophical practice. Rejecting this move and sticking with the more realistic notion of prima facie conceivability, we are now going to inspect how these flaws based on our fallible human (psychological) condition threaten the reliability of the faculty of imagination and thus undermine the basis of our modal intuitions.

5.2.3. Prima Facie Conceivability is Unrealiable

From the definition of the two versions of conceivability (prima facie and ideal) arises one immediate problem: Whether something will be prima facie conceivable for someone is a matter of how much thought, how much cognitive resources she invests. This problem arises because ideal conceivability is a rational concept that rules out any contradiction already by definition, while prima facie conceivability is a psychological notion and is thus subject to human flaws and errors.

As I have argued, (Chalmers, 2002) simply assumes ideal conceivability and does not attempt to defend the merits of the more modest prima facie version. When defending his Zombie argument in his earlier book *The Conscious Mind* however, Chalmers argues that there is no substantial reason to think that cognitive limitation poses a real threat to the faculty of imagination. (Chalmers, 1996) He frames the objection by looking at the seemingly apt analogy with the necessity of certain complex mathematical truths: Mathematical hypotheses like the Goldbach conjecture, he says, seem to constitute counterevidence for the claim that mere conceivability is a good guide to possibility. This is because as they are still unproven, both their truth as well as their falsity is conceivable; and because either their truth or their falsity is necessary only one option is possible. It follows that one of these options is impossible but still conceivable.

Countering the threat of these mathematical hypotheses to the link between conceivability and possibility, Chalmers argues that in one important respect the analogy between mathematical and philosophical intuition goes astray: While in the mathematical case our modal reasoning leaves matters open, in the zombie case we have a strong intuition in favor of the possibility of a zombie world. His verdict then is to give modal intuition the benefit of doubt: "While it must be conceded that any philosophical argument *could* go wrong because of cognitive impairment, in the absence of any substantial reasons to believe this, this sort of objection seems quite *ad hoc*" (Chalmers, 1996, p. 139f).

I am partly sympathetic to Chalmers' reply. On the one hand, I share Chalmers' overall sentiment via the objection: *En gros*, the modal intuitions involved in the method of cases seem to be unproblematic, so an opponent of that method should do more than simply point to the mere possibility of cognitive impairment. To be fair, it has to be noted that the cognitive limitation objection *does* try to be more specific, namely by limiting its application to "complex" scenarios. Still, it is unclear how to delineate that subgroup of problematic "complex" cases, so on first view Chalmers' rebuttal seems to stand true.

On the other hand, I do not believe that the analogy with the necessity of certain complex mathematical truths is so bad. I don't think that in the philosophical case the mere strength of a modal intuition can sufficiently ground

the possibility of the scenario under consideration. An objector could rehabilitate the analogy with mathematics by looking at clashing intuitions in the world of philosophy on an interpersonal, rather than an intrapersonal level. Here, the philosophical literature in general and the zombie debate in particular provide us with evidence of just these kinds of diametrically opposed strong intuitions. Provided that these clashing intuitions are really about the same content, they present a formidable analogy to the mathematical case. As such, the objection from cognitive limitation remains strong.

Cognitive limitation is not the only problem the method of cases has to deal with. In the following part, we will see how underdetermination exists not only at the level of the description of possible worlds but also at the conceptual level, which further diminishes the usefulness of the method of cases as a tool for conceptual clarification.

6. Holding on to the Method of Cases

Despite all my concerns about Parfit's cases and Chalmers' attempt to ground possibility in ideal positive imagination, the idea of strong intuitions being evidence on their own remains plausible to me. Contrasting the mostly unanimous reaction of philosophers in the Gettier case with the mixed reactions in the zombie case might hint towards a possible way to handle the situation. One

way to think about the problem of misleading intuitions is to ask: What do the seemingly problematic cases have in common?

A very common answer to this question is the often voiced concern that some thought experiments are "mere science fiction", or in other words too outlandish, too esoteric, too far-fetched to judge reliably. As an example of this claim, take the following passage from a quite recent experimental philosophy paper by Jonathan Weinberg:

> "Intuitions may be fine as a class, taken on the whole, and the opponent has neither the need nor the desire to attack that whole class. But [...] the practice appears to set no constraints on how esoteric, unusual, far-fetched, or generally outlandish any given case may be. [...] So this anything-goes aspect of the practice is what makes it particularly ripe for the opponents' challenge." – (Weinberg, 2007, p. 321)

One thing that speaks against linking inconceivability to outlandishness and conceivability to ordinariness is the fact that those characteristics don't always go together so easily. In response to the above cited passage, Herman Cappelen points out that the *esoteric*, *unusual* and *far-fetched* cases are not in any way correlated to cases which are difficult to judge. To illustrate that point, Cappelen brings up Perry's and Burge's thought experiments (Perry, 1979; Burge, 2007), which feature events that happen all the time – Burge even explicitly stresses the ordinariness of his cases –, but are hard to judge.

As an easily judicable but far-fetched case Cappelen cites the following:

> "Suppose there are two pink elephants in my office. Then yet another pink elephant comes into my office (and the first two pink elephants stay in the room). Question: How many pink elephants are in my office?" – (Cappelen, 2012, p. 226)

You could also modify the case so that it is just the opposite from the case above, i. e. easily imaginable but very hard to judge:

> Suppose a pink elephant is walking past you. As it gets to the other side of the room and touches the wall, it suddenly vanishes from your sight. Do you *know* that a pink elephant walked through the room?

This case is seemingly easy to imagine (*pictured*) in the sense of having a modal intuition, but very hard to judge in the sense of having a proper conceptual intuition. Many concerns cloud our faculty of imagination in this case, concerns like: Is it still an elephant if it is pink? Should the fact that what you saw so clearly contradicts the laws of physics point to some epistemic failure on your side, undermining your justification/reliability?

These two cases illustrate that outlandishness is not easily related to inconceivability. With this criterion ruled out and no other good replacement to be found in the literature, I consider the question of how to delineate good from bad cases of (prima facie negative) imagination still an open question, the discussion of which would exceed what I hope to accomplish for my thesis.

Part II: Conceptual intuition and indeterminate concepts

7. Visiting Twin Earth

Putnam's Twin Earth example is often cited as one of the most influential thought experiments in the history of philosophy. This fictional story about a planet not much different from our own has become what Gabriel Segal called a "sort of paradigm in the philosophies of language and mind" (Segal, 2000, p. 24). Its publication challenged the received view of meaning, which held that (1) the meaning that a speaker associates with a word is determined by individualistic facts about that speaker, and (2) the meaning of a word determines its extension. Putnam's thought experiment convinced many philosophers that those two assumptions cannot be jointly satisfied. His rendition of the thought-experimental scenario spans the better part of three pages, so I shall only quote it in abbreviated form (Putnam, 1975, p. 139f):

> "For the purpose of the following science-fiction examples, we shall suppose that somewhere in the galaxy there is a planet we shall call Twin Earth. (...) [A]part from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is exactly like Earth. (...) One of the peculiarities of Twin Earth is that the liquid called 'water' is not H_2O but a different liquid whose chemical formula is very long and complicated. I shall abbreviate this chemical formula simply as XYZ. I shall suppose that XYZ is indistinguishable from water at normal temperatures and pressures. In particular, it tastes like water and it quenches thirst like water. Also, I shall suppose that oceans and lakes and seas on Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water etc.

> If a spaceship from Earth ever visits Twin Earth, then the supposition at first will be that 'water' has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that 'water' on Twin Earth is XYZ, and the Earthian spaceship will report somewhat as follows:
>
> 'On Twin Earth the word "water" means XYZ'"

Now imagine that an inhabitant of planet Earth named Oscar and an inhabitant of Twin Earth named $Oscar_2$ both think about water. Putnam crafts his fictional story in a way to bring out that we intuitively judge Oscar and $Oscar_2$ to have different thought contents, meaning that they think about different stuff.

Remember that the received view of meaning held that (1) the meaning that a speaker associates with a word is determined by individualistic facts about that speaker, and (2) the meaning of a word determines its extension. As Putnam wants to hold on to the connection between meaning and extension (claim 2), his move is to say that individualistic facts alone cannot determine meaning (negating claim 1). The difference in thought contents can only be explained with reference to the extension of the natural kind – i. e. H_2O in Oscar's case, and XYZ in $Oscar_2$'s case. In other words: The content of a word is no longer determined by individualistic facts about the speaker (*narrow content*), but is also sensitive to the external world (*wide content*). "Cut the pie any way you like", he famously concluded, "'meanings' just ain't in the head!" (Putnam, 1975, p. 144)

Instead of deriving his conclusion from some background theory, Putnam cites his intuition as (counter)evidence against the received view of meaning. He

then proposes a different theory, which can account for the intuition. According to one characterization by Jerry Fodor, the Twin-Earth Problem "isn't a problem; it's just a handful of intuitions together with a commentary on some immediate implications of accepting them" (Fodor, 1987, p. 208).

In the opening sentence of the section, where Putnam tells the Twin Earth story, he announces that his claim will be "shown with the aid of a little science fiction." (Putnam, 1975, p. 139) Instead of first telling the story and then giving us his intuition on it, he weaves the intuition cleverly into the narrative. The key passage is the following one:

> If a spaceship from Earth ever visits Twin Earth, then the supposition at first will be that "water" has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that "water" on Twin Earth is XYZ, and the Earthian spaceship will report somewhat as follows:
>
> "On Twin Earth the word 'water' means XYZ." (Putnam, 1975, p. 140)

Or to make Putnam's intuition explicit: When uttered on planet Earth, the word 'water' refers to H_2O, but when uttered on Twin Earth it refers to XYZ.

7.1. The Debate that Followed

The implications that Putnam's thought experiment had were huge: it touched central tenants of the philosophy of mind, the philosophy of language and also epistemology. And it took the philosophical world by surprise. As Paul Boghossian described his reaction at a recent conference on intuition:

> "[T]hat I was tempted to make that verdict having read through the thought experiment came as a big surprise to me. I tried to resist it, but it kept forcing itself back upon me. It seemed like an unexpected and significant new realization. I despise it even to this day. It has made a lot of trouble. But it can't be helped." (Boghossian, 2013)

Many philosophers shared that feeling and quickly went on to draw out the conclusions that followed from Putnam's results, without paying attention to the fact that the intuition is not unanimously accepted. In his critique of the method of reflective equilibrium, Robert Cummins criticizes the way philosophers reacted to the Twin Earth results:

> "It is commonplace for researchers in the Theory of Content to proceed as if the relevant intuitions were undisputed. Nor is the reason for this practice far to seek. The Putnamian take on these cases is widely enough shared to allow for a range of thriving intramural sports among believers. Those who do not share the intuition are simply not invited to the games. This kind of selection allows things to move forward, but it has its price. Since most nonphilosophers do not share the intuition, the resulting theories of content have little weight with them, and this is surely a drawback for a theory that is supposed to form an essential part of the foundations of cognitive psychology." (Cummins, 1998, p. 116)

Cummins only speaks of nonphilosophers diverging from the Putnam intuition, but professional philosophers were divided on this issue just as well. Apart from Cummins himself, the whole group of philosophers ascribing to a view called *descriptivism*, shares his intuition on the Twin Earth stories, including philosophers such as John Searle, Hugh Mellor, and Timothy Crane.

7.1.1. Water or No Water? Descriptivism vs. Realism

In a recount of the debate that followed Putnam's Twin Earth experiment, Barry Barnes frames these diverging intuitions in terms of a disagreement between descriptivists and realists. Descriptivists hold the view that "the extension of a kind term is fixed by a verbal specification of a set of manifest properties" (Barnes, 1982), in other words: what a term means is fixed by its definition alone. This would be the view of John Searle, Hugh Mellor or Timothy Crane (Searle, 1958; Crane, 1991; Mellor, 1977). Realists, on the other hand, claim that when a term is first applied to a particular thing or instance, it 'baptizes' (or 'christens' or 'dubs') that thing or instance. This theory is also called the causal theory of reference because the term finds its way into the linguistic community through dissemination, which is a causal process. Putnam, of course, is a proponent of this view, but also for example Saul Kripke and Tyler Burge are on this side of the intuitive divide. (Putnam, 1975; Burge, 2007; Kripke, 1980)

In his discussion of the debate, Barnes pits Putnam's realist intuition against Mellor's descriptivist intuition:

> "Putnam suggests that the new material should be set without the extension of 'water' because it has a different microstructure [...] Mellor, in contrast, sees nothing objectionable in the descriptivist alternative of holding that water has been discovered to vary in its microstructure." (Barnes, 1982, p. 30f)

Barnes goes on to draw a wide-reaching metaphilosophical conclusion based on the form of the debate, which we will discuss later. Before we do that, I want to cite one further reaction to the Putnam intuition, which can't be subsumed under the two camps that Barnes listed above.

7.1.2. Questioning the Scientific Basis of Twin Earth

Trying to save his concept of incommensurability in the face of rigid designation (which would be able to fix a sample's reference even across scientific revolutions), Thomas Kuhn offers his own interpretation of the Twin Earth story (Kuhn, 1990).

Putnam describes Twin Earth as a place that "apart from the differences we shall specify in our science-fiction examples … is *exactly* like Earth." (Putnam, 1975, p. 139) Said difference is that on Twin Earth the liquid called water is composed of a substance with a long, complicated formula, abbreviated as XYZ. It is a substance that, as Putnam stresses, "is indistinguishable from water at normal temperatures and pressures" (Putnam, 1975, p. 140). (We shall put aside the point – noted by Stalnaker (1993) and others – that Twin Oscar cannot truly be identical to Oscar, given the fact that the human body mostly consists of H_2O.)

But while Putnam describes the scientist who visits Twin Earth as judging the watery stuff there not to be 'water', Kuhn gives a different description of what would happen: The report that visitors send home about the stuff that lies in

Twin Earth's lakes "should not be about language but about chemistry", he writes. "It must take some form like: 'Back to the drawing board! Something is badly wrong with chemical theory.'" (Kuhn, 1990, p. 310)

Now one might interpret Kuhn as simply offering another diverging intuition. After all, he gives us one more story of what the scientist "would say". On this view, we would have three different intuitions corresponding to three things the scientist might say: "that stuff is not 'water'" (Putnam and others), "that stuff is 'water'" (Mellor and others), "something is badly wrong with chemical theory" (Kuhn and others). However, even though the debate is indeed often framed in terms of asking what the scientist visiting Twin Earth would say, I think this way of putting the question is misleading. After all, Putnam wants to find out whether the extensions of the two linguistic communities' word 'water' are overlapping or non-overlapping. It might very well be that the visiting scientist would react by saying: "We need to rewrite all of our chemistry", but even then Putnam might *still* ask whether or not the estranged visitor would refer to the stuff on Twin Earth as 'water'. The compatibility of the visitor's reaction with this question shows that the third possible reaction isn't the manifestation of a third competing intuition on the same question, but rather an artifact of posing the question in an imprecise manner. The better way to phrase the question, then, is to ask: If Oscar on Earth and Twin Oscar on Twin Earth utter (or think) the word 'water', do they mean the same thing? This is a question to which Kuhn doesn't really give an answer. Therefore I

think Kuhn is best understood as questioning the scientific basis of the thought experiment:

> "The terms 'XYZ' and 'H_2O' are drawn from modern chemical theory, and that theory is incompatible with the existence of a substance with properties very nearly the same as water but described by an elaborate chemical formula. Such a substance would … demonstrate the presence of fundamental errors in the chemical theory that gives meanings to compound names like 'H_2O' and the unabbreviated form of 'XYZ'." (Kuhn, 1990, p. 310)

Recent scientifically more comprehensive support for this position comes from Christopher Grisdale. He points out that our chemistry tells us that there is no possible world which is (1) exactly like ours but where at the same time (2) watery stuff is not H_2O. This is because water's microstructure significantly influences its macrostructure. As Paul Thagard summarizes Grisdale's work:

> "even a slight change in the chemical constitution of water produces dramatic changes in its effects. If normal hydrogen is replaced by different isotopes, deuterium or tritium, the water molecule markedly changes its chemical properties. Life would be impossible if H_2O were replaced by heavy water, D_2O or T_2O; and compounds made of elements different from hydrogen and oxygen would be even more different in their properties." (Thagard, 2012)

To sum up this third reaction to the Twin Earth case: Because our current scientific theories tell us that there is no possible world where something conceptually identical to our concept of water is physically different from H_2O, the stipulated scenario has nothing to do with "water" and thus we gain no relevant insights from testing our intuition here.

In the following section we will try to understand the limitations of our concepts and what they mean for the method of cases.

8. Underdetermination of Concept Extension

Decades before Twin Earth or zombies or other strange stories entered the philosophical world, there was the occasional philosopher that cautioned us about the shortcomings of our concepts in the face of unusual borderline cases.

Ludwig Wittgenstein suggests that the method of cases might not be suitable for situations, which are so far from our actual world, that our concepts do not fit these circumstances.

> „‚Es ist als wären unsere Begriffe bedingt durch ein Gerüst von Tatsachen.‘
>
> Das hieße doch: Wenn du dir gewisse Tatsachen anders denkst, sie anders beschreibst, als sie sind, dann kannst du die Anwendung gewisser Begriffe dir nicht mehr vorstellen, weil die Regeln ihrer Anwendung kein Analogon unter den neuen Umständen haben. – Was ich sage, kommt also darauf hinaus: Ein Gesetz wird für Menschen gegeben, und ein Jurist mag wohl fähig sein, Konsequenzen für jeden Fall zu ziehen, der ihm gewöhnlich vorkommt, das Gesetz hat also offenbar seine Verwendung, einen Sinn. Trotzdem aber setzt seine Gültigkeit allerlei voraus; und wenn das Wesen, welches er zu richten hat, ganz vom gewöhnlichen Menschen abweicht, dann wird z. B. die Entscheidung, ob er eine Tat mit böser Absicht begangen hat, nicht etwa schwer, sondern (einfach) unmöglich werden.“ (Wittgenstein, 1967, Z. 350)

In a very similar vein, W. V. O. Quine wonders whether language is cut out for such extraordinary circumstances:

> "The method of imaginary cases has its uses in philosophy, but at points [...] I wonder whether the limits of the method are properly heeded. To seek what is 'logically required' for sameness of person under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with." (Quine, 1972, p. 490)

His background assumption is that language is an instrument that was developed to describe our actual world, and this is where it can be used successfully. In situations that are so extraordinary that they don't occur in our everyday life at all, language consequently fails to be an adequate tool for description.

In *Reasons and Persons*, Parfit addresses this objection:

> "Quine's and Wittgenstein's criticism might be justified if, when considering such imagined cases, we had no reactions. But these cases arouse in most of us strong beliefs." (Parfit, 1986, p. 200)

Parfit is probably right to notice that these cases evoke strong beliefs in most of us. However, leaving aside that empirical matter, simply pointing to such strong reaction cannot ground the epistemic value of our intuitions. After all, we have seen that intuitions can and do clash, and from analogy with perception we know that even in cases where our judgments systematically align, like in the case of optical illusions, they might be misleading us. By explaining their faulty etiology, we may defuse the intuition even if not curing us of it. Barnes provides us with such an explanation. So, as promised earlier, we now come to the metaphilosophical conclusion that Barnes drew from examining the Twin Earth debate.

8.1. Meaning Finitism

Barnes recounts the Twin Earth debate in terms of two opposing camps that each proposed a theory of meaning of their own and relied heavily upon examples of normal accepted usage, or modification of usage. These two camps cited their respective intuitions as support while at the same time acknowledged (but failed to account for) a number of counterexamples. When it came to hypothetical situations, he says, "both sides [were] able to gloss them to their own satisfaction." (Barnes, 1982, p. 30)

Even though Barnes does not exactly explain how each of the clashing intuitions came to be, he gives an explanation of why it is possible for them to clash in the first place. Barnes' answer to that riddle is a theory of concepts known as *meaning finitism*. According to that view, concepts do not have a fixed extension. The application of a concept is not fully determined by its definition; it is rather a matter of contingent judgment by the actors of a language community.

Central to language learning under meaning finitism is the concept of exemplars. As a young child gets exposed to instances of cats, she acquires an *array of exemplars* of what her language community will accept as falling under the concept of a 'cat'. Some of those exemplars will be strongly paradigmatic, bearing more importance to classifying instances under that concept, while others

will be less central or even borderline cases. The child thus develops a linguistic disposition to apply the concept to certain individuals while withholding application in other instances. Judgments about concept application thus become judgments of similarity to more or less paradigmatic instances of a concept.

Meaning finitism does away with clear-cut definitions and sees concepts as social institutions

> "[T]here is no utility in the notion of the extension of a concept ... Far from the meaning of a concept fixing its future proper use, we can now see that people judge how to develop the use of a concept, and that imputations of meaning can do no better than to follow on behind, rationalizing the effects of sequences of such judgments." (Barnes, 1982, p. 32)

Because of this rejection of this understand of concepts as social institutions, meaning finitism also rejects what Martin Kusch called the three central tenants of the orthodox philosophy of meaning: semantic determination, the notion of fixed, unchanging extensions, and the central role of truth in semantics.

In the original paper from 1982 Barnes presents his conclusions as if they followed easily from the phenomenon of two clashing intuitions in a single thought experiment on the extension of 'water'. Apart from that observation, he offers no substantive argument in favor of a finitist semantics, or, conversely, against meaning determinism. However, regardless of this shortcom-

ing in his original paper from 1982, the underlying idea of a language community making up, shaping and constantly reshaping the extension of our concepts seems to be so apparent as to be almost undeniable. Still, this concession to the fleeting nature of our concepts is seldomly reflected in actual philosophical discourse, and so "all too often at present we adopt a finitist approach when studying knowledge and an extensional approach when celebrating it." (Barnes, 1982, p. 38)

Barnes would later go on to develop a program called the *sociology of scientific knowledge* (SSK) together with David Bloor and John Henry. In their programmatic book "Scientific Knowledge: A Sociological Analysis" (Barnes, Bloor, & Henry, 1996) the authors further develop the meaning finitist view and summarize their position in five main theses (ibid, pp. 55–59):

(1) "the future applications of concepts are open-ended": Because concept application is based on judgments of similarity and because the array of exemplars is different for every speaker and shifts also intra-personally, the extension of a concept cannot be fixed at any point in time.

(2) "no act of classification is ever indefeasibly correct": The dichotomy of correct and incorrect is the product of a social institution, depending on the consensus of a language community, and is thus always subject to change: "People must *decide* what is correct and what is not." (ibid, p. 56)

(3) "all acts of classification are revisable": With acts of classification never being indefeasibly correct, and always being subject to change, a community might revise certain acts of classification. An example that has received much attention from the media is the re-classification of Pluto, which is no longer a planet.

(4) "successive applications of a kind term are not independent": Because each application of a kind term influences the array of exemplars, successive judgments of similarity are influenced by earlier uses of a term.

(5) "the applications of different kind terms are not independent of each other": Concepts hang together in complex ways, and, accordingly, concept applications are interdependent too. How some individuals use 'duck' may affect how others use 'goose'.

In a sense, this view echoes Quine's position on the method of cases, only that it is thought to its radical conclusion. The problem with the method of cases, to repeat the gist of Quine's view, is the false assumption that "words have some logical force beyond what our past needs have invested them with" (Quine, 1972, p. 490). While Quine's remark is situated in the context of the personal identity debate, with examples arguably far removed from our everyday situations, Barnes takes Quine's idea one step further and extends it to each and every act of concept application:

> Concept application is always a matter of contingent judgment in every particular case. No act of concept application is ever fixed or determined by previous acts of concept application or by alleged 'meanings' intrinsic to concepts. (Barnes, 1982, p. 33)

I cannot here delve into the pros and cons of meaning finitism as much as I'd like to. For discussions of its benefits, I refer the reader to (SSK1996, ch. 3; Bloor, 1997; Barnes, 1992). For a discussion of its problems see Ilkka Niiniluoto's section on finitism in his book "Critical Scientific Realism" (Niiniluoto, 2002, pp. 262-268).

I will conclude this section by drawing out the consequences that meaning finitism, if true, has for the method of cases and the deliverances of intuition, and link the theory with recent work in epistemology and the philosophy of intuition.

9. Rethinking Intuition and the Method of Cases

As I see it, the implications of meaning finitism for the method of cases are two-fold: First, the contingency of concept application implies that diverging intuitions can and must be explained in psychological and sociological terms. My judgment over a matter of concept application reflects both facts about general human psychology and individual interests. Insofar as the relevant part of my psychology or of my interests is widely (or even universally) shared,

my judgment will be more likely to fall in line with other people's judgments. Jennifer Nagel makes use of an empirical theory of intuition (cf. my section on Nagel) to explain the dialectical strength of intuitions. Consensus is not necessarily the same as correctness, however, and an intuition might be widely shared and still be misleading. Tamar Gendler examines these kinds of *intuitive illusions* (cf. the section on Gendler), and shows how the structure of the concept, which the thought experiment is intended to explore, is a big influence on whether or not the thought experiment can succeed. As far as sociological analyses of concept application go, there is, of course, the very coarse-grained work of the experimental philosophers, but these studies are mostly quantitative and seem content with showing basic regularities correlating to sociological categories. If meaning finitism is taken seriously, illuminating the philosophical debates over intuitive responses to thought experiments could be an interesting field of research, which is still largely neglected. Early examples of this kind of approach can be found in David Bloor's "Knowledge and Social Imagery" (Bloor, 1976), Steven Shapin's and Simon Schaffer's "Leviathan and the Air-Pump" (Shapin & Schaffer, 1985), and Martin Kusch's book on "Psychologism: A Case Study in the Sociology of Philosophical Knowledge" (Kusch, 1995).

Second, the meaning finitist view poses a severe challenge to the method of cases. In their reliance on thought experiments philosophers show a mostly uncritical implicit commitment to meaning determinism (and probably also to

the classic theory of concepts). As Martin Kusch observed, most of mainstream philosophy of language falls within the 'meaning-determinist' camp (Kusch, 2005, p. 201) – and I suspect this is true for the vast majority of philosophers throughout all the subfields of philosophy. However, the status of the method of cases as an objective philosophical method is incompatible with the claims of meaning finitism. Following that view, the faculty of intuition loses its privileged epistemic status as a bearer of objective conceptual truth, because concepts are inherently open-ended, and their future application is "always a matter of contingent judgment in every particular case" (Barnes, 1982, p. 33).

This does not mean that these judgments are random. A concept's meaning is to a large part determined by its being accepted by a linguistic community and is thus a social category. Personal interests also play a role, but only insofar as they are shared by some proportion of the surrounding linguistic community. Shared interests are much more powerful determinants of conceptual judgments, "because they enter in many more acts of judgments and because they lead to collective actions." (Kusch, 2005, p. 206) My intuitive judgment of a particular situation should thus be understood as a manifestation of my linguistic disposition and how I personally *wish* the concept to be extended; it simply reflects my community's and my individual interests. In a recent book on Epistemology, Richard Foley develops a new approach to the definition of knowledge, where he brings in the interests and judgments of speakers in a

way very reminiscent of the teachings of meaning finitism, a relationship to which I devote another section.

9.1. Jennifer Nagel

Even though Gettier's intuition ran counter to all established theories of knowledge, he felt that his intuition would be shared by other people as well and that it was so strong that it would be convincing without further argument. Examples like these show that intuition is more than a "1-bit signal: is p possible, yes or no?", as Jonathan Weinberg put it (Weinberg, 2007, p. 335). It carries information also about the extent to which one's own response will be stable and will be shared by other people when confronted with the same situation. However, how is that possible? How can we know, how others would respond to a certain situation?

Drawing from Asher Koriat's work in the field of psychology, Jennifer Nagel applies the Self-Consistency Model (SCM) (Koriat, 2008; Koriat, 2012) to epistemic intuition. Koriat's SCM predicts a subject's response and confidence in two-alternative forced-choice questions. When faced with such a question, the subject will sample representations from a pool (think: an array of paradigms) in order to produce a response. Each representation will shift the balance in one direction or the other, and the size of the sample's majority will determine the confidence that is associated with the answer. Koriat found that a subject's confidence in an intuitive judgments predicts both the likelihood that she will

make that same judgment on subsequent occasions, and that other people will share her judgment. This is predicted by the model because high confidence means that the majority of the sample is internally consistent, which – on balance – will be reproducible on subsequent samplings both by the same persons and by other people.

If a subject samples from a pool of representations that are mostly in favor of a ‘yes’-answer, then chances are very high that the sample will contain representations principally in favor of a ‘yes’, thus yielding a ‘yes’-answer with very high confidence. If the sample is more balanced, the subject might still answer ‘yes’, but with only little confidence.

In that way, the strength of an intuition (i. e. one’s confidence) will predict the likelihood of making the same choice on subsequent occasions. Moreover, if other people are sampling from a similar pool of representations, then the strength of an intuition will also predict *their* answers. Thus, one’s confidence in a response predicts both the stability of one’s own response across subsequent applications as well as consensus across a group of people sharing a similar array of representations.

This latter point helps explain the immense dialectical success of philosophical intuition. Because an individual’s intuitive response is indicative of how others will respond when presented with the same story, we are able to conduct most (if not all) of our philosophical enterprise from the armchair.

Consensus is not correctness, however. It is perfectly conceivable that there is intuitive consensus on a question without the answer being correct, "intuitive illusions" similar to perceptual ones. In her work, Tamar Gendler looks at such imaginary cases gone wrong.

9.2. Tamar Gendler

As we have heard above, meaning finitism holds that meaning is a social construct, subject to change. People must decide what is correct, and these decisions are always revisable. When people make these judgments, they are not only influenced by their linguistic dispositions, but also by human psychology. In a paper titled "Exceptional Persons: On the Limits of Imaginary Cases" (Gendler, 1998) Tamar Gendler shows how subtly these judgments can be led astray.

She argues that the structure of the concept that the thought experiment is intended to explore has a significant influence on how informative the thought experiment can be. Using two cases from Bernard Williams, she argues that the concepts of personhood and personal identity are not organized around necessary and sufficient conditions, but rather through the continued coincidence of enough of the factors that ordinarily allow us to persist over time. We judge the fringe cases only by "cantilevering out from the set of generally-obtaining correlations which characterize ordinary cases." (Gendler, 1998, p. 608) In other words, the fringe cases are exemplars by courtesy only.

In a paper called "Personal Identity and Thought-Experiments" (Gendler & Hawthorne, 2002) Gendler expands on that idea and explains this cantilevering process in value judgments. Many thought experiments make use of what John Stuart Mill called the *method of agreement*, which holds that "[i]f two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon" (Mill, 1973, p. 390). That is what Gendler called the exception-as-scalpel strategy in her earlier paper. Gendler argues that this principle, useful as it is for causal explanations, can mislead us in certain cases where we want to explain value judgments. Gendler calls these cases instances of *borrowed lustre*, "where both pure and impure instances of a phenomenon are accorded the same assessment because impure instances are treated as relevantly similar to pure ones." (Gendler & Hawthorne, 2002, p. 47)

To illustrate this, she gives examples of both an unproblematic as well as a problematic assessment:

> "Suppose that whenever I strike a match against the side of a matchbox and say 'Let there be light', the match bursts into flame; whenever I strike a match against the side of a matchbox and say nothing, the match bursts into flame; whenever I simply hold the match in the air and say 'Let there be light', the match remains unlit; and whenever I neither strike the match nor recite the incantation, the match remains unlit." (Gendler & Hawthorne, 2002, p. 42)

Which gives:

	Striking match against box	No striking against box
Let there be light	Flame	No flame
[silence]	Flame	No flame

Using the method of agreement here correctly shows that it is the striking against the box, rather than the verbal utterance, which ignites the match.

Gendler contrasts this with problematic cases like the following:

> "Suppose we venerate regular geometrical figures for their beauty, but certain approximations to regular figures also produce the same respect by way of resembling the ideal. We might portray the circumstances as follows:" (Gendler & Hawthorne, 2002, p. 47)

	X is square-like	X is not square-like
X is a square	X is an appropriate target of veneration	
X is not a square	X is an appropriate target of veneration	X is not an appropriate target of veneration

The chart reveals that whenever something is square-like, it is an appropriate target of geometrical veneration, and whenever something is not square-like, it is not an appropriate target of geometrical veneration. The method of agreement fails in cases like these: It would be mistaken to think that it is square-likeness rather than proper squareness that explains the appropriateness of geometrical veneration. What explains the veneration is rather the approximation to an ideal, i. e. (the approximation to) ideal squareness.

> "The reason for this is that in borrowed-lustre cases one of the antecedent conditions for application of the method is not satisfied: the way in which A brings about P in the A-plus-C case is different from the way A brings about P in the A-not-C case. When I venerate a perfect square, its square-like features (A) cause me to venerate it (P) because of the irresemblance to a feature that it has, squareness; whereas when I venerate a merely approximate square, its square-like features (A) cause me to venerate it (P) because of their resemblance to a feature that it lacks, squareness." (Gendler & Hawthorne, 2002, p. 47)

If Jennifer Nagel showed how our intuitions are indicative of the extent to which they are shared with other people, the caveat of intuition is that it can be subject to illusions the same way as perception is. With examples like the borrowed lustre case, Tamar Gendler makes a first step towards identifying the subtle ways in which commonly shared intuitions can lead us astray. However, in the same way as perceptual illusions don't undermine our faculties of perception wholesale, it would be unwarranted to dismiss intuition completely as a philosophical tool simply because it can go astray in certain situations.

9.3. Richard Foley

In a recently published book called "When is True Belief Knowledge?" (Foley, 2012), Richard Foley suggests a new approach to the analysis of knowledge. His approach fits neatly with the meaning finitist view outlined above.

Traditional proposals for the analysis of knowledge all assume that what needs to be added to justified true belief is something related to, but distinct from

true belief. Some traditions seek it in a special kind of justification (non-defective, indefeasible, …), others try to qualify the process and faculties that produce or sustain a belief (reliably generated, truth-tracking, the product of properly functioning cognitive faculties, …). Foley suggests that what is in fact needed to get from true belief to knowledge are *more true beliefs*. Not any true beliefs, however, but those which are deemed important in the evaluative context. Restating the problem as one of important information allows Foley to zoom out of the problem space and better account for the full diversity of those instances we call knowledge: "Although there is a variety of such shortcomings, it can be tempting to fasten upon stories involving a particular kind of shortcoming and to try to build an entire theory of knowledge around them." (Foley, 2012, p. 22)

By making the test of what's important relative both to the situation and the concerns, values, and interests of the community judging that situation, he can elegantly subsume all those competing analyses of knowledge under his account: "[J]ustification theorists, reliability theorists, or proponents of other approaches … provide a directory to the sorts of gaps that are apt to strike observers as important." (Foley, 2012, p. 24)

Foley's view fits neatly with the idea of contingent concept application. Remember that this position holds that concepts are malleable rather than fixed, and each and every concept application amounts to a contingent judgment on

the part of an actor in a language community. We can recognize this fundamental point of meaning finitism in Foley's account too: Whether or not the concept of 'knowledge' applies to a certain situation is not inherently contained in some fixed definition of knowledge, but is determined by important information – information which depends on contingent facts about the actor, her concerns, values and interests.

Conclusion and Outlook

In this thesis, we looked at the impact of underdetermination on the philosophical method of cases. Our first insight was that this method employs two kinds of intuition: first we use a modal intuition, judging some stipulated case to be possible; only then we employ a conceptual intuition, judging a given case to be an exemplar of the concept under examination. These two kinds of intuition both come with certain epistemological limitations that owe themselves to the inherent underdetermination of our imagination in the former case and of our concepts in the latter case. I dedicated one part of my thesis to each of these two kinds of intuition and their respective limitations.

❧

In the first part of my thesis we visited Parfit's combined spectrum case to see how conceivability can lead us astray, even if it may sustain prima facie consideration. We then looked at Chalmers' theory of modal rationalism, which attempts to provide an epistemic link between conceivability and possibility. I argued that Chalmers fails in his attempt because he needs ideal positive conceivability for it to be a good guide for possibility. My argument went like this:

1. positive conceivability is so vaguely defined that it fails to be distinguishable in practice from negative conceivability;
2. ideal conceivability is irrelevant for actual philosophical practice; so

3. we are left with prima facie negative conceivability, which is too unreliable to be a good guide to metaphysical possibility.

In case of doubt then, a modal intuition is not a good guide to possibility. I ended this first part by looking at some attempts to cordon off the bad from the good cases of conceivability, noting that any such attempt to date has been wanting. However, there still is a pre-theoretic feeling that there is *something* different about the Gettier cases on one side and the Twin Earth and zombie cases on the other side, even if it has not been properly explicated yet. With this feeling in the vicinity, there is hope that future research may be able to work out helpful criteria to separate the wheat from the chaff and thus lay a better foundation for the epistemological reliability of our thought experiments.

❧

For my second part, we looked at Putnam's Twin Earth cases, a storyline so popular in modern philosophical discourse that Gabriel Segal called it a "sort of paradigm in the philosophies of language and mind". We heard Barnes' framing of the Twin Earth debate as a clash of intuitions between followers of the realist and the descriptivist camps of meaning. Following Barnes further we arrived at the theory called meaning finitism, which can account for both the largely undisputed conceptual intuitions and also for the clashing intuitions in borderline cases. Meaning finitism holds that concepts do not have a

fixed extension, and as such the application of a concept is not fully determined by its definition, it is rather a matter of contingent judgment by the actors of a language community.

I suggested that the theory of meaning finitism explains both why the largely homogenous group of western analytic philosophers could rely on intuition for so long until the first experimental philosophers stepped outside of that group and found diverging intuitions co-varying with all kinds of social categories – or as I would stress: sociolinguistic categories. This realization brought us to the scientific program called the *sociology of scientific knowledge* (SSK) that Barry Barnes worked out together with David Bloor and John Henry. Here we found new insights about the malleability and social character of our concepts, leading us toward a revised theory of intuition and a more modest epistemological claim for the method of cases, which can only gain insights into concepts relative to some language community.

Wrapping up this second part, we looked at a few authors who have written articles and books that fit very well with this new sociolinguistic setting of our conceptual intuition. Jennifer Nagel provides an empirical theory of intuition to explain the dialectical strength of intuitions. Consensus is not necessarily the same as correctness, however, and an intuition might be widely shared but still misleading. Tamar Gendler examines these kinds of intuitive illusions and

shows how the structure of the concept, which the thought experiment is intended to explore, is a big influence on whether or not the experiment can succeed. We found the most fruitful overlap in Richard Foley's recent book "When is True Belief Knowledge?" Foley's epistemological treatise suggests that what is really needed to get from true belief to knowledge are more true beliefs. Not any true beliefs, however, but those which are deemed important in the evaluative context. Whether or not the concept of 'knowledge' applies to a certain situation is not inherently contained in some fixed definition of knowledge, but is determined by important information – information that depends on contingent facts about the actor, her concerns, values, and interests. Notice that importance is always relative to an actor or a group of actors, so through the concept of "important information" we again recognize the necessity for a sociological relativization of the goal of the method of cases, conceptual analysis.

❧

My thesis situated the limitations of intuition on two levels: the level of imagination and modal intuition; and the level of conceptual intuition. As I laid out in the chapter on conceivability and possibility, I consider the question of how to delineate good from bad cases of (prima facie negative) imagination still an open question. For my outlook, I want to list briefly the three possible direc-

tions one could investigate in order to find an answer to that question, borrowed from Tamar Gendler and John Hawthorne's book *Conceivability and Possibility* (Gendler & Hawthorne, 2002, p. 10):

The first strategy for separating good and bad cases of conception might be on the basis of subject matter. Perhaps there are certain types of propositions (abstract metaphysical ones, ones concerning necessary beings, ones that turn on actual empirical matters of fact, etc.) that are ill-suited for conceivability arguments.

The second strategy is based on defining the right kinds of conceiving. Perhaps it is only clear and distinct conceivings or conceivings accompanied by rational insight, or maybe ones that involve a detailed intellectual vision of a possible scenario, which can get to the status of an ideal conception, i. e. a conception, for which no contradiction is detectable even on ideal rational reflection. I have discussed both Descartes' and Chalmer's attempts to ground possibility in clear and distinct modal intuitions above, and as said before, I don't see much promise in this strategy.

Third, one could also combine the previous two strategies, either by connecting the certain types of subject matter to certain types of conceiving or by restricting one or both of these domains to some unproblematic subdomain. Only further research will show, which of these strategies will prove to be the most fruitful.

When it comes to the level of conceptual intuition, I believe that meaning finitism explains both the findings of experimental philosophers *and* the relative stability of intuitions that allowed us to be blind to the intuitive disagreements for so long. Re-interpreting the empirical findings through the lens of the sociology of scientific knowledge would bring a fresh and – I believe – more realistic view of what is going on in these empirical and theoretical cases.

For the method of cases, it first and foremost means being more modest, because any claim to objectivity is no longer sustainable. The method remains a useful tool for sounding our philosophical concepts, but we must not forget that these findings are always relative to a certain (linguistic) system of concepts. There remains a lot to be said about philosophers' latent commitment to an outdated view of concepts, because, to repeat Barnes' dictum, "all too often at present we adopt a finitist approach when studying knowledge and an extensional approach when celebrating it." (Barnes, 1982, p. 38) A glance over to the works of linguists, most importantly Ferdinand de Saussure's structuralist approach, may help us understand the diachronic and synchronic dynamics of how a community's concepts hang together and shift over time. As far as I know, today's experimental philosophers are still only concerned with more of a static view of concepts. Future research could bring out diachronic shifts that are to be expected for the borderline cases of our most contested concepts.

Appendix

References

Adams, R. M. (1989). Should Ethics be More Impersonal? A Critical Notice of Derek Parfit, Reasons and Persons. *The Philosophical Review*, *98*(4), 439–484.

Barnes, B. (1982). On the extensions of concepts and the growth of knowledge. *The Sociological Review*, *30*(1), 23–44. doi:10.1111/j.1467-954X.1982.tb00652.x

Barnes, B. (1992). Realism, relativism and finitism. In D. Raven, L. Van Vucht Tijssen & J. de Wolf (Eds.), *Cognitive Relativism and Social Science* (pp. 131–47).

Barnes, B., Bloor, D. & Henry, J. (1996). *Scientific Knowledge: A Sociological Analysis*. Chicago and London: Chicago University Press and Athlone Press.

Bealer, G. (1998). Intuition and the Autonomy of Philosophy. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry* (pp. 201–239). Lanham: Rowman & Littlefield Publishers.

Bealer, G. (2004). The Origins of Modal Error. *Dialectica*, *58*(1), 11–42.

Bloor, D. (1976). *Knowledge and Social Imagery*. Chicago: University of Chicago Press.

Bloor, D. (1997). *Wittgenstein, Rules and Institutions*. London: Routledge.

Boghossian, P. (2013). Talk on Herman Cappelen's Philosophy Without Intuition. London.

Burge, T. (2007). Individualism and the Mental. In *Foundations of Mind: Philosophical Essays, Vol. 2* (Vol. 4, pp. 100–181). Oxford: Oxford University Press.

Cappelen, H. (2012). *Philosophy Without Intuitions*. OUP Oxford.

Chalmers, D. J. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Chalmers, D. J. (2002). Does Conceivability Entail Possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 145–200). Oxford: Oxford University Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Cohnitz, D. (2005). *Gedankenexperimente in der Philosophie* (Vol. 7163). Paderborn: Mentis.

Crane, T. (1991). All The Difference in the World. *The Philosophical Quarterly*, *41*(162), 1–25.

Cummins, R. C. (1998). Reflection on Reflective Equilibrium. In M. DePaul & W. Ramsey (Eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry* (pp. 113–127). Lanham: Rowman & Littlefield.

Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting* (Vol. 1984, p. 200). MIT Press.

Descartes, R., Cottingham, J., Stoothoff, R. & Murdoch, D. (1984). *The Philosophical Writings of Descartes: Volume II*. Cambridge: Cambridge University Press.

Deutsch, M. (2010). Intuitions, Counter-Examples, and Experimental Philosophy. *Review Of Philosophy and Psychology*, *1*(3), 447–460.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*.

Foley, R. (2012). *When Is True Belief Knowledge?*. Princeton and Oxford: Princeton University Press.

Gendler, T. S. (1998). Exceptional Persons: On the Limits of Imaginary Cases. *Journal Of Consciousness Studies*, *5*(5-6), 592–610.

Gendler, T. S. & Hawthorne, J. (2002). Introduction. In *Conceivability and Possibility* (pp. 1–70). Oxford: Oxford University Press.

Gettier, E. L. (1963). Is Justified True Belief Knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. I. (2007). Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status. *Grazer Philosophische Studien*, *74*(1), 1–26.

Hume, D. (2000). *A Treatise of Human Nature*. (D. F. Norton & M. Norton, Eds.). Oxford: Oxford University Press.

Knobe, J. & Nichols, S. (2007). An Experimental Philosophy Manifesto. In *Experimental Philosophy*. Oxford University Press.

Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(4), 945–59. doi:10.1037/0278-7393.34.4.945

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80–113. doi:10.1037/a0025648

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Kuhn, T. S. (1990). Dubbing and redubbing: The vulnerability of rigid designation. In C. W. Savage (Ed.), *Minnesota Studies in the Philosophy of Science* (Vol. 14, pp. 298–318). Minneapolis: University of Minnesota Press.

Kusch, M. (1995). *Psychologism: A Case Study in the Sociology of Philosophical Knowledge*. London: Routledge.

Kusch, M. (2005). *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford: Oxford University Press.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal Of Philosophy*, *61*(4), 343–377.

Mach, E. (1897). Über Gedankenexperimente. *Zeitschrift für den physikalischen und chemischen Unterricht*, *10*, 1–5.

McGinn, C. (1989). Can We Solve the Mind-Body Problem? *Mind*, *98*(391), 349–366.

Mellor, D. H. (1977). Natural Kinds. *The British Journal for the Philosophy of Science*, *28*, 299–312. doi:10.1007/s10441-008-9056-7

Mill, J. S. (1973). A System of Logic, III viii 2. In J. M. Robson (Ed.), *Collected Works, Vol. VII*. University of Toronto Press.

Niiniluoto, I. (2002). *Critical Scientific Realism*. New York: Oxford University Press. doi:10.1093/0199251614.001.0001

Ørsted, H. C. (1822). Über Geist und Studium der allgemeinen Naturlehre. *Gehlens Journal für Chemie und Physik*, *36*(4), 458–488.

Parfit, D. (1986). *Reasons and Persons* (p. 560). Oxford: Oxford University Press.

Perry, J. (1979). The Problem of the Essential Indexical. *Noûs*, *13*(December), 3–21.

Pust, J. (2012). Intuition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter2012 ed.). Retrieved from http://plato.stanford.edu/entries/intuition/

Putnam, H. (1975). The Meaning of "Meaning". *Minnesota Studies In the Philosophy of Science*, *7*, 131–193.

Quine, W. V. O. (1972). Review of Identity and Individuation. *The Journal Of Philosophy*, *69*(16), 488–497.

Searle, J. R. (1958). Proper Names. *Mind*, *67*, 166–173.

Segal, G. M. A. (2000). *A Slim Book About Narrow Content*. Cambridge: MIT Press.

Shapin, S. & Schaffer, S. (1985). *Leviathan and the Air-Pump*. Princeton: Princeton University Press.

Shoemaker, S. (1985). Reasons and Persons. By Derek Parfit. *Mind*, *94*(375), 443–453.

Simpson, P. (1985). Reasons and Persons by D. Parfit. *The Review Of Metaphysics*, *39*(2), 370–372.

Slezak, P. (2013). *Content Externalism and Intuition*.

Sosa, E. (2007). Intuitions. In *A Virtue Epistemology* (pp. 44–69). Oxford: Clarendon Press.

Stalnaker, R. (1993). Twin Earth Revisited. *Proceedings Of the Aristotelian Society*, *93*, 297–311.

Stalnaker, R. (2002). What is it Like to be a Zombie? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 385–400). Oxford: Oxford University Press.

Thagard, P. (2012). *The Cognitive Science of Science*. MIT Press.

Weinberg, J. M. (2007). How to Challenge Intuitions Empirically Without Risking Skepticism. *Midwest Studies In Philosophy*, *31*(1), 318–343. doi:10.1111/j.1475-4975.2007.00157.x

Williamson, T. (2007). *The Philosophy of Philosophy*. New York: Routledge.

Wittgenstein, L., Anscombe, G. & von Wright, G. H. (1967). *Zettel*. University of California Press.

Wolf, S. (1986). Self-Interest and Interest in Selves. *Ethics*, *96*(4), 704–720.

Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy And Phenomenological Research*, *53*(1), 1–42.

Abstract

This master's thesis examines the impact of underdetermination on the philosophical method of cases. One finding is that this method employs two kinds of intuition: first philosophers use a modal intuition, judging some stipulated case to be possible; only then they employ a conceptual intuition, judging a given case to be an exemplar of the concept under examination. These two kinds of intuition both come with certain epistemological limitations that owe themselves to the inherent underdetermination of our imagination in the former case and of our concepts in the latter case.

Zusammenfassung

Die vorliegende Masterarbeit beschäftigt sich mit Rolle und Auswirkung der inhärenten Unterdeterminiertheit jedes Gedankenexperiments zur Klärung philosophischer Konzepte. Meine These lautet, dass diese Methode der Begriffsklärung auf zwei Arten von Intuitionen beruht: Zunächst bedient man sich einer modalen Intuition, um zu etablieren, dass eine fingierte Situation auch tatsächlich möglich ist; darauf aufbauend erst befragt man seine Intuition hinsichtlich des zu prüfenden Konzeptes. Beide Arten der Intuition, so die These, sind von epistemologischen Einschränkungen geprägt, die sich der inhärenten Unterdeterminiertheit sowohl unserer Vorstellungskraft – im Falle der modalen Intuition – als auch unserer begrifflichen Konzepte verschulden.

Curriculum Vitae

Alexander Wolfgang Seifert von Eichenstark, BSc BA

* 16.07.1985 in Wien

Akademische Laufbahn

- 1991–1995 Volksschule (Sacré Coeur Wien)
- 1995–2003 Gymnasium (Theresianische Akademie Wien)
- 2003–2010 BSc Software & Information Engineering (TU Wien)
- 2008–2011 BA Linguistik, Zweig Grammatiktheorie (Uni Wien)
- 2011–2015 MA Wissenschaftsgeschichte und Wissenschaftsphilosophie (Uni Wien)
- 2011–2012 Tutor beim Eingangskolloquium *Methoden und Probleme der Wissenschaftsphilosophie und Wissenschaftsgeschichte*
- 2012–2013: Wissenschaftlicher Mitarbeiter beim ERC-Projekt *Distortions of Normativity*