Doctoral Thesis

Constructivism without Verificationism

Casper Storm Hansen

Candidatus Magisterii
Københavns Universitet

A thesis presented for the degree of Doctor of Philosophy at the University of Aberdeen

2014

University of Aberdeen

Contents

Declaration of authorship **iv**

Summary **v**

Preface **vii**

Introduction **2**

1 Infinity **6**
- 1.1 The iterative conception of sets . . . 8
- 1.2 Transfinite ordinals . . . 18
- 1.3 The continuum hypothesis . . . 21
- 1.4 Countable infinity . . . 22
- 1.5 Intuitionism . . . 32

2 Free choice sequences **38**
- 2.1 Free choice sequences according to Brouwer . . . 38
- 2.2 Constitution of free choice sequences . . . 39
- 2.3 Main argument . . . 46
- 2.4 Definite descriptions . . . 48
- 2.5 Future contingents . . . 49
- 2.6 Fictional objects . . . 54
- 2.7 Consequences for the continuum . . . 56

3 Brouwer's conception of truth **61**
- 3.1 An equivalence in propositional logic . . . 72
- 3.2 The proof of the Bar Theorem . . . 74
- 3.3 Ordering axioms . . . 79
- 3.4 Two truths versus BHK . . . 80
- 3.5 Lawless choice sequences revisited . . . 81

4 Non-verificationist constructivism **84**
- 4.1 Dummett's argument from meaning constraints . . . 85
- 4.2 Truth as potentiality . . . 92
- 4.3 Wittgenstein's rule following considerations . . . 95
- 4.4 Arithmetic . . . 104

5 Classes and real numbers **115**
5.1 Sets, classes and species . . . 115
5.2 Kripke's theory of truth . . . 119
5.3 A theory of classes and real numbers . . . 124
5.4 Unrestricted comprehension . . . 131
5.5 The basics of classes and sequences . . . 133
5.6 Real numbers . . . 135
5.7 Expressive weakness . . . 137
5.8 Bishop's diagonal proof . . . 139
5.9 Diagonalization . . . 141
5.10 Comparison with Skolem's paradox . . . 149

6 Compositional supervaluation **151**
6.1 Gupta's Challenge . . . 152
6.2 Grounding . . . 154
6.3 Kripke's supervaluation . . . 156
6.4 Revision theory . . . 158
6.5 The alternative . . . 161
6.6 The formal theory . . . 165
6.7 Meeting Gupta's Challenge . . . 170
6.8 A generalization . . . 172
6.9 Theorems and proofs . . . 174
6.10 Comparison and discussion . . . 181

7 Supervaluation on posets **185**
7.1 Introduction and examples . . . 185
7.2 Considering all possibilities . . . 196
7.3 ...but not too many . . . 199
7.4 The formal theory . . . 201
7.5 Expressive strength . . . 206
7.6 Consistency . . . 213
7.7 TIC and TAP . . . 214
7.8 Conventionalism and failure of bivalence . . . 216

Conclusion **220**

References **223**

Declaration of authorship

I, Casper Storm Hansen, confirm that the work presented in this thesis, entitled "Constructivism without Verificationism", is my own. I declare that it only contains quotations that have been clearly distinguished as such by quotation marks or indention and attributed to the source from which they came. Furthermore, this work has not been accepted in any previous application for a degree.

Signed: Casper Storm Hansen

Date: July 3, 2014

Summary

Historically, two claims have been correlated in the philosophy of mathematics. The first is that the objects of mathematics are what can be constructed in time by a subject. The second is that a mathematical sentence can be true only if it is proved (or provable). In this dissertation it is argued that these claims are independent of one another and that, in fact, the former is correct and the latter incorrect. This opens up the possibility of a mathematics that is closer to classical mathematics than intuitionism is, even though it is based on the austere ontological basis of mental constructions. I lay the groundwork for such a mathematics.

It is first argued that the reasons for being skeptical towards actual infinity are so strong that mathematics should not be based on it; it is much more unclear what is in the "Cantorian paradise" than normally assumed, and supertasks (including a new one presented here) imply absurd consequences of actual infinity. Instead, we will have to make do with mental constructions and potential infinity. In that respect this dissertation sides with the intuitionists.

However, my position is far from that of the intuitionists. Among other things, I argue that choice sequences do not have the properties that Brouwer claims. They do not in themselves lead to failure of bivalence. And lawless sequences implicitly depend on actual infinity and therefore cannot do the job for the theory of real numbers he assigns them.

I give a novel interpretation of Brouwer in terms of two notions of truth: a narrow one called "truth-in-content" and a more liberal one called "truth-as-anticipation", that are conflated in the so-called Brouwer-Heyting-Kolmogorov interpretation. On that background it is argued that Brouwer's rejection of verification-transcendent truth, and acceptance of *tertium non datur* for decidable but undecided propositions, is incoherent in its shifting appeal to in some cases actual and in other cases potential mental constructs.

A foundation for a stronger mathematics is provided by allowing an even more liberal notion, "truth-as-potentiality", defined consistently in terms of potential mental constructs. Dummetian and Wittgensteinian objections to it are discussed and rejected. In this way, it is shown that it is possible to avoid the damaging consequences of intuitionism: the necessity of rejecting large parts of classical mathematics, that there seems to be overwhelmingly good reasons for believing are in order as they are. In particular, classical arithmetic is vindicated on this new basis.

So mentalism does not lead *by itself* to revision of logic. On the other hand, I argue that a mentalistic theory of collections can be, and, in so far as we want it to be comprehensive, has to be, non-well-founded and that *this* forces a revision of logic. However, the result is not intuitionistic logic but something akin to the logic that comes out of Kripke's theory of truth: *tertium non datur* fails only in the absence of groundedness, not in the absence of proof and decidability.

I mix Kripke's theory with elements of Bishop's constructive analysis to build a non-verficationist constructivist theory of classes and real numbers and show that in this setting Cantor's diagonal proof does not prove that there are more real numbers than there are natural numbers, but that the real numbers are, in a certain sense, indefinitely extensible.

The "akin to" reservation is due to the fact that there are serious problems with the specific formal theory given by Kripke, in particular problems with expressing universal generalizations about the object language itself and with a certain family of problematic sentences pointed out by Gupta. Being sympathetic to the general direction of his proposal, I therefore extend and modify the theory. These technical tweaks I justify through a careful analysis of the notion of "grounding"; where Kripke only gives it metaphorical motivation, it is here interpreted as temporal dependency – some mental constructs can only be effected after certain other mental constructs. The notion of truth-as-potentiality plays a crucial role here. In particular, it can be used to get rid of the reliance on classical transfinite ordinals that Kripke uses in his formulation of the theory.

Preface

The night before my oral high school mathematics exam, preparing for the topic *infinity*, my doubts, which had lingered for some time, about what I had been told by my teacher about Hilbert's Hotel and Cantor's theory of the infinite and was now expected to explain back to her, reached a point where I could no longer ignore them. I happened to know a PhD student in mathematics, whom I therefore called and asked a few critical questions. He assured me that the theory of the transfinite was on a solid footing. And I accepted it. Mathematics is absolutely certain, I knew, so if the mathematicians agree that there are higher infinities, then that is how it is.

It was therefore with great anticipation that I a few years later started on a course on logic and set theory as part of my minor in mathematics at the University of Copenhagen. Now I would get the full story, and I expected to be convinced. When there was virtually universal agreement among mathematicians, it had to be because they had very strong reasons that it was merely a matter for me to take in. Deep was my disappointment with the mere metaphors and the reliance on the power of three dots that I found their case to rest on.

Over the following years I gradually formed my own ideas on the subject. The philosophy education in Copenhagen was flexible and allowed me the time to delve deep into any subject I fancied writing a paper on. I took a first stab at the believer in actual infinity, learning a lot in the process, although I did not at the time realize just how challenging an opponent she is. A new course in the philosophy of mathematics, that was introduced at just the time I needed it, also gave me the opportunity to engage seriously with intuitionism, the rebellious opposition to main stream mathematics that I had until then only had a vague impression of. Although finding the idea fascinating, it was clear to me that it was far too extreme. It seemed incomprehensible to me that no one had explored the position that I perceived right between classical mathematics and intuitionism.

At around the same time an instructor pointed me towards Hofstader's *Gödel, Escher, Bach*, a work that, although I could not make myself agree with its conclusions, kindled my imagination in a way that no other book has done before or after. I also found Kripke's *Outline of a Theory of Truth*, which at the same time pointed in the right direction, I was and am convinced, and left a lot of improvement work for me to do. I spent hundreds of hours lost to the surrounding world and with my eyes focused to a point at infinity, contemplating the mysteries and paradoxes of self-reference. Together the two texts set me on the course to independent and original contributions to philosophy. A visit to the Institute of Logic, Language and Computation in Amsterdam provided me with the skills to develop my ideas in a rigorous and clear way, allowing me to write my master thesis and two publications on my first ideas on how to generalize and modify Kripke's theory.

This dissertation is the provisional culmination on that intellectual journey. I believe I have managed to develop my previously embryonic idea into a comprehensive and coherent theory on the nature of mathematics, taking the golden middle path between classical mathematics and intuitionism.

I have written the dissertation during two academic years spent at the Northern Institute of Philosophy at the University of Aberdeen and one academic year visiting the Plurals, Predicates and Paradox research group at the University of Oslo.

I am indebted to a number of people who helped me through these three years and made them very interesting. My foremost thanks goes to my primary supervisor, Crispin Wright, the sharpest thinker I have had the privilege of knowing.

Øystein Linnebo, my secondary supervisor and host during the Oslo stay, generously offered his time and insight whenever I needed it and could be counted on for critical but always constructive feedback. To him and the other core members of his research group, Jon Litland, Jönne Kriener and Sam Roberts, I am grateful for many stimulating discussions in and outside of seminars.

Aaron Cotnoir also acted as a secondary supervisor in my first year and helped me improve on the first rough drafts of Chapters 5 and 6. Toby Meadows gave Chapters 5 and 7 a final check.

Mark van Atten met with me on my two visits to Paris to discuss drafts of Chapters 2 and 3. He forcefully defended Brouwer against my attacks, thus

– I hope – ensuring that those attacks are now properly directed at Brouwer and not at a straw man. The second of those two Paris trips was due to the kind invitation of Michael Detlefsen. Thanks to him I spent the entire month of June last year in the charming capital of the French, attending logic talks and living well.

My office mate, Andreas Fjellstad, made sure it was never boring to come to work and was kind enough, in between discussions of *tonk*, to leave me occasional moments to work on this dissertation. To the remaining members of NIP, past and present, and the rest of the philosophy department I can only say that I am sad to leave you and that you will always hold a special place in my heart.

A very special thanks goes to my parents, Marianne and Jesper, who provided me with both a generous loan, without which this endeavor would not have been possible, and a welcoming home in Denmark to return to on holidays.

Finally, I should, in the interest of full disclosure, acknowledge a debt to my own former self: a handful of paragraphs in sections 5.2, 6.3 and 6.4 are more or less copy-pasted from my master's thesis.

Casper Storm Hansen
Aberdeen, July 2014

Constructivism without Verificationism

Introduction

The primary problem in the philosophy of mathematics is to find out what it is we are talking about when we talk about mathematics. L.E.J. Brouwer gave a wonderfully simple answer to that question, namely that we talk about the structure of our mental experiences. It is a wonderfully simple answer that has been almost universally rejected; not so much because of its intrinsic implausibility, but rather due to the consequences for logic he drew from that ontological thesis, which turned the beautiful queen of the sciences, believed to live in a paradise, into a horrid witch expelled to a dark underworld.

The ontological thesis is correct, I believe. The consequences Brouwer drew from it are largely mistaken. The actual consequences are much less detrimental. Building from mental constructions alone, we can erect a mathematics that is much closer (in so far as a metric for such an abstract space can be assumed to be given) to classical mathematics than intuitionism is. That is the agenda for the present dissertation.

Formulated with a little more precision, the central task of the philosopher of mathematics is to make ontological, semantical and epistemological sense of mathematics, i.e. to locate a part or aspect of Being that can serve as the subject matter of mathematics and provide referents for the mathematical terms and truth makers for the theorems in such a way that the fact that we have mathematical knowledge can be explained. But which mathematics? There are classical mathematics, intuitionist mathematics, other types of constructivist mathematics, strict finitist mathematics, etc. Which mathematics is it that the philosopher should make sense of? In spite of the revisionism I will champion in this work, I think that it is classical mathematics for which ontological, semantical and epistemological justification should *first* be attempted. Classical mathematics is both the simplest and the most powerful mathematics to work with, so if there is a philosophical foundation for it to be found, doing so would be the greatest accomplishment possible in this branch of philosophy.

Let us pretend, just with the aim of making a point, that there were some particular part of Being that all mathematicians took to be the subject matter of their science and with reference to which they practiced classical mathematics. Even in that counter-factual situation it would not suffice for the opponent of classical mathematics to argue that *this* part of Being did not justify classical mathematics. For if there were some other part of Being that did, then the mathematicians would be vindicated (at least *qua* mathematicians, if not as metaphysicians) and could go on doing what they did. (In that way, it seems, mathematics is different from every other science.) As it happens, there is no agreement about what (classical) mathematics is aiming to describe. But that does not detract from the point: only if there is *no* possible foundation for classical mathematics is classical mathematics wrong. And if it is wrong, we should find a philosophical foundation with which we can salvage as much as possible, without any prejudice about where to look.

It will be argued in Chapter 1 that we have established no foundation for classical mathematics and that there is room for serious doubt about its possibility. On that basis we will reject classical mathematics as a mathematics that has the level of epistemic certainty that we should require of something we rely so heavily on in the sciences as we do with mathematics. Then we will go in search of the ontologically acceptable mathematics that is closest. Although we will conclude that it is among mental constructions that we shall locate mathematical objects, Chapters 2 and 3 and parts of Chapter 4 are concerned with showing that intuitionism is not it. Constructivism is accepted, verificationism is rejected.

The main characters in this dissertation will be the *Platonist* and the *intuitionist* and then the so-called *non-verificationist constructivist*, whose part I will play myself, and who will dominate the stage from Chapter 4 onwards. The *strict finitist*, who believes that infinity has no legitimate role to play, not even in its potential form, and the *mathematical empiricist* will be minor characters. There are a couple of characters that one might expect to find in this play, who do not appear (outside of this paragraph). One is the *logicist*. The reason is that the most fundamental question we will be concerned with is whether mathematical propositions are about physical, mental or abstract entities, and the logicist is not concerned with that. He claims that mathematical propositions are logical propositions, but that just pushes the question back to what it is that makes purely logical propositions true or false. The *formalist* will also not play a role, except for the purpose of the occasional comparison. She will not because our conclusion will imply that such desperation is not called for. The *structuralist* will not be admitted to the scene because he responds to a perceived luxury problem,[1] namely that there are several different structures in Being that can do the work of, e.g., the classical real numbers. The conclusion in this work will be that there are probably none. Our challenge will be to find structures that exemplify the kind of structures that classical mathematics deals with or, failing that, some that come "as close as possible". The problems that may arise in case several different structures for the same theory can be located are insignificant in comparison and will be set aside. At the end of Section 1.1, the Platonist will have evolved into a modal Platonist and she is a sort of *fictionalist*.

Chapter 2 is about Brouwer's lawless choice sequences and will be entirely negative, arguing that they do not have the salient properties attributed to them by Brouwer and that they do not add to the theory of real numbers in the way he claims. Chapter 3 gives a novel interpretation of Brouwer's conception of truth. The so-called Brouwer-Heyting-Kolmogorov interpretation is rejected *qua* exegetical thesis about Brouwer's intuitionism and replaced by an interpretation operating with both a strong notion of truth, "truth-in-content", and a weaker but intimately connected, "truth-as-anticipation". In Chapter 4 I will again argue against Brouwer, but here the aim is not purely negative. I will not argue that the two notions of truth are illegitimate. Separating constructivism and verificationism, I will just make the point that Brouwer is mistaken about the claim that truth-as-anticipation is the weakest legitimate notion of truth. I will claim that that honor befalls an even weaker type of truth, "truth-as-potentiality", which is also "grounded" in truth-in-content. On the basis of that conception of truth and thereby an ontology of mental constructions, classical arithmetic will be vindicated.

The rest of the dissertation is concerned with theories of collections, and here the conclusion is not as positive: classical set theory can (unsurprisingly) not be salvaged. We will reject ZFC together with its underlying idea, that of the combinatorial notion of collections. Following Brouwer we will turn instead to the logical notion, where a collection is characterized by a criterion of membership. However, we will not follow him very far. There is (surprisingly) a connection between constructivism and Kripke's theory of truth, and the logic that is used in that theory is significantly closer to the bulls-eye than intuitionistic logic. That will be the subject of Chapter 5, where we will investigate the fate of real numbers and diagonalization in such a theory of classes. We

[1] The structuralist's slogan, in the case of arithmetic, is "any ω-sequence will do" (see, e.g., (Hellman 1989) and (Shapiro 1997)). I agree. I just think, for reasons to follow, that the supply of ω-sequences is very meager. Shapiro's *ante rem* structuralism is Platonism in disguise (see in particular pages 72 and 93-94) and is thus affected by the critique in Section 1.1. Hellman's modal structuralism survives that far but is subject to the problems considered in Sections 1.2-1.4.

will also point out the problems of expressive weakness that arise in Kripke's theory of truth and which exist in analogous form in the class theory that results from transferring Kripke's principles directly. In Chapter 6 a first stab is taken at solving those problems by proposing an alternative way to do supervaluation, which, unlike Kripke's own, only assigns the value of true or false when that value is genuinely grounded at that point in the construction. This first attempt is educational but ultimately unsatisfactory. Therefore, a second theory with the same aim is developed in the final chapter.

Chapter 1

Infinity

The triviality of the sentence "7 is larger than 6" as seen from an intra-mathematical point of view contrasts sharply with the significance of the philosophical problems of explaining the semantics of that sentence and accounting ontologically for its truth. It has the same surface structure as "Mt. Everest is larger than Kilimanjaro" which refers to two objects and is true because a certain relation obtains between them. So the desire for uniformity of our theory of semantics compels us to locate referents for "7" and "6" somewhere in Being and explain the truth of the mathematical sentence on a par with the truth of the geographical one. And given that, Ockham's Razor further demands of us that we, at least, make an attempt at doing so without making stipulations about what is in Being that we would not otherwise have made.

The (mathematical) empiricism of John Stuart Mill (1895) is such an attempt. The numeral "7" is claimed to denote aggregates of seven physical objects. And the sentence "7 is larger than 6" is taken to express the fact that any aggregate of seven physical objects encompasses an aggregate of six physical objects. In general, mathematical propositions are claimed to be nothing but highly general laws of nature. So, given that we already accept physical objects and aggregates thereof in our ontology, a semantics for this sentence and sentences like it has been secured for free. Semantic uniformity and metaphysical austerity are achieved to a degree that no other philosophical theory of mathematics can compete with.

The problems with mathematical empiricism are, however, both obvious and well-known. The thesis implies that if physical objects suddenly changed their behavior so that every time a group of two objects were placed next to another group of two objects, two of those objects would merge into a single object, it would become true that 2 plus 2 equals 3. But we are not open to possible revisions of mathematics on that sort of basis. Mathematical propositions may aid us in describing the physical world, but only in the hypothetical way that *if* some physical system satisfies certain mathematical axioms *then* it satisfies the consequences of those axioms. For instance, *if* pebbles do not merge or multiply when placed next to each other, so that pebbles satisfy the axioms of Peano arithmetic, *then* placing 2 pebbles next to 2 pebbles will result in a collection of 4 pebbles. The mathematical propositions are not themselves about the physical world and do not depend on it for their meaningfulness or truth values – at least not in this simple-minded way.

Another unacceptable consequence of mathematical empiricism is that if there are only 10^{80} atoms in the universe, as the astro-physicists estimate, then there is no referent for the term "10^{100}", nor a meaning or a truth value for sentences containing it. It is obviously both meaningful and true to say that 10^{20} times 10^{80} equals 10^{100}. And even if we were to stipulate that such a proposition is to be interpreted as an assertion about the physical world, and hence to be counted as not true, it would be very hard to deny that the intended content of that proposition can instead be expressed both meaningfully and truthfully by the counterfactual "if there had been a sufficient number of atoms in the universe, then 10^{20} times 10^{80} would equal 10^{100}". But empiricism cannot support this, short of being amended with a story about reference to other possible worlds and thereby being morphed into a completely different theory.

Mathematical empiricism can only supply the ontological and semantical underpinnings for a strictly finite mathematics.[1] But at least some measure of infinity must be admitted into mathematics, or one is left with a very difficult task of explaining away the success and apparent meaningfulness of, e.g., mathematical analysis.

Hence the appeal of Platonism; a realm of abstract mathematical objects, existing independently of the physical and mental world(s), is postulated to secure the meaningfulness of mathematical terms and to supply truth makers for what is standardly taken to be mathematical truths.[2] If we insist on understanding mathematical sentences on a par with other sentences of similar grammatical structure, it may seem *necessary* to postulate such extra entities,

[1]The limitation to the finite would be lifted if arbitrary regions of space-time were allowed as objects and there are infinitely many such regions. However, there would still be a problem with more complex mathematical entities such as for example the set of all real functions. Kitcher (1998) tries to solve this problem, and vindicate Mill, by using iterated constructions of aggregates of physical objects. He thereby turns the empiricism into a form of constructivism where it is the subjective acts of collecting that really serve as truth makers, and the "physical urelemente" are reduced to a role which, on the same assumptions, could just as well be played by more subjective acts of collecting.
[2]Some classical references are (Bernays 1935) and (Gödel 1947).

so doing so is sanctioned by Ockham's Razor, which states that "plurality is not to be posited *without necessity*".[3]

1.1 The iterative conception of sets

Nevertheless, subscribing to Platonism is a high price to pay metaphysically. Worse, if you are in the market for a justification of contemporary mathematics and pay that steep price, you do not get what you were bargaining for. Slightly simplified, contemporary mathematics is what follows from ZFC by classical logic, and Platonism does not support ZFC. So I will argue. In this section we will consider different suggestions for how Platonism could do that job, beginning naïvely and gradually examining more sophisticated options, but ultimately finding them all lacking. In Sections 1.2 to 1.4 we will put further pressure on classical mathematics. The chapter closes with a section introducing intuitionism.

To Plato himself, some of the most important characteristics of the World of Ideas are that it is eternal and unchanging and that it is determinate.[4] If we follow Plato in also assuming that the World of Ideas contains all mathematical objects, and (anachronistically) let that include sets, then these characteristics seem to lead to paradox. The World of Ideas presently contains all possible sets and it is determinate for each set whether it contains itself. So the sets that do not contain themselves are sharply distinguished from those that do and therefore they should form a set in the World of Ideas, for the World of Ideas contains *all* the sets. But, of course, the set of all the sets that do not contain themselves can neither contain itself nor not contain itself, but by determinacy it must either contain itself or not contain itself.

Contemporary mathematics is not subject to such Russellian[5] treats of inconsistency. The reason is that ZFC is based on the iterative conception of sets.[6] According to this doctrine, the sets are exactly the things that are created in a transfinite process of discrete stages, where at each stage each plurality of already existing sets forms a new set.[7] So starting with no sets, we form all the possible sets of existing sets, which is just one, the empty set, $\varnothing$. Having that one set, there are now two different pluralities of sets, namely the empty plurality and the plurality encompassing just the empty set, and therefore at

[3]"[P]luralitas non est ponenda sine necessitate" (de Ockham 1979, 322).
[4]See in particular *The Republic*, *Meno* and *Parmenides* in (Plato 1997).
[5]See (Russell 1902).
[6]Or rather, that is *one* way to motivate ZFC. I will focus on it as I think it is the most reasonable. For exposition and defense of the iterative conception see (Boolos 1971), (Wang 1974, chapter 6) and (Parsons 1975).
[7]As has become standard, I am here ignoring the possibility of *urelemente*.

the next instance we can create two sets, $\varnothing$ and $\{\varnothing\}$. Those two sets make for four possible pluralities of sets, so at the next stage there are four sets. And so on through all finite stages. After all those stages, there is the first transfinite stage where a countable infinity of sets is available to form an uncountable infinity of new sets. And so on forever and ever...

Implicit in the above derivation of the Russellian contradiction was the assumption of full comprehension: for any property there is a set of all and only those sets that have the property. The iterative conception of sets gives a story that can be used as a background for denying full comprehension. For it is not in general the case that a property is such that there is a stage at which all the sets with that property have been formed, so as to be available to form the intended set. Only restricted comprehension is valid in the iterative universe: for any given set x and any property there is a set of all and only those sets in x that have the property. In particular there is no stage where all the sets that do not contain themselves (which according to the iterative conception are *all* the sets) are already formed, so there is no Russell Set. (Thus, it is claimed, Russell's Paradox never threatened the concept of set, but at most the concept of classes. Focus on this distinction is postponed to Section 5.1.)

So Russell's Paradox is avoided – but only by making inconsistent assumptions! On the one hand, the World of Ideas is supposed static, and on the other, the explanation of the iterative conception of sets is full of dynamic expressions such as "create", "process", "already existing", "form", "next instance" and "and so on forever and ever".[8] Having freed ourselves of the first inconsistency, we must now seek liberation from the second. Let us consider possible ways out within a Platonic setting.

One possibility is to fully accept the iterative conception and reject the notion that the World of Ideas is static. After all, unchangeability is not part of the definition of "Platonism" as specified above, namely as the thesis that abstract (mathematical) objects exist independently of the physical and mental,[9] and there is no reason to feel bound by the teachings of Plato just because we are using his name. So *prima facie* it is a theoretical option that the Platonic universe of sets might be changing. We could, without backing away from the goal of justifying the iterative conception, compromise with Plato and say that

[8]Gödel, in his (1947), seems to endorse the view I am here accusing of being inconsistent.

[9]In fact one could argue that it does follow directly from the definition: time is an aspect of the physical/mental world, so if what exists in the Platonic world is relative to a time index, the existence of the abstract entities is not independent of the physical world, contrary to the definition. But in itself this observation of a minimal form of dependence is hardly a major problem for the Platonist, just a cause for minor modification of the exact formulation of the thesis.

it is only changing by expanding, thus holding on to the claim of the eternity of each Platonic entity forward in time.

In spite of this being the direction in which the currently widely accepted notion of sets seems to push the Platonist, no one, to my knowledge, has endorsed it.[10] And for good reason. If the Platonic realm is dynamic there is some specific state it is in right now. What could possibly determine which one? What is the rate of change? And what is the force that causes these changes? It seems absurd to believe in a thesis that leads to such questions.[11]

However, that was nothing but an appeal to intuition through rhetorical questions, and I should do better. So here is another stab: the semantics is not as intended. A universal generalization over all sets that is true today, may tomorrow be false. For then the sentence may be about more sets and a counterexample may have been created overnight. More generally, sentences that are theorems of ZFC are at times false. There was a time when it was false that there is a set of all real numbers. Also, the proposition that every set is an element of some other set, remains false. For at any given time there is a stage which is the most recent. This follows from the iterative conception: to form the sets of a given stage, its elements must already be available. Only one level of sets can be formed at any given time.

For this reason, it would also seem that the axiom of infinity remains false. If known physical discrete stage processes can be taken as a model, this set creating process would up till now only have gone through a finite number of stages and therefore only a finite number of sets would have been created. To avoid that consequence the Platonist would have to appeal to supertasks[12], making further heavy metaphysical assumptions on top of those already inherent in Platonism. It is, as far as known, logically possible that an infinity of stages have been completed, even if the set formation at each stage takes a positive amount of time. It "just" requires that there is no lower bound on that duration. The set of the first stage could be created in 1/2 second, the sets of the second stage in 1/4 second, the sets of the third stage in 1/8 second, and so on. That way, the first ω_0 stages are run through in just 1 second.

However, the invocation of supertasks would not take the Platonist very far. Even if supertasks, the completion of infinitely many distinct, temporally disjoint tasks in finite time, are possible, hypertasks, the completion of *uncountably* many distinct, temporally disjoint tasks in finite time, are not. This was

[10] Maddy (1990) thinks that the universe of sets changes because when physical objects are created or destroyed, ur-elemente are created or destroyed and with them the impure sets they give rise to. This kind of change is not relevant to our present discussion.

[11] See (Parsons 1975) for critique of temporal and quasi-temporal interpretations of the iterative hierarchy.

[12] The classical papers are (Thomson 1954) and (Benacerraf 1962).

argued in (Clark and Read 1984), using the assumption that any task takes positive time: Given any origin of a time scale and unit of time, and in addition a well-ordering of the rational numbers, there will for each task be a first rational instant of time in its period of duration. By temporal disjointness, the function taking each task to that corresponding instant of time will be injective. Therefore, as the image of the function, a subset of the set of rational numbers, is countable, so is the domain.

In the present context the force of the argument can be strengthened, for the assumption is superfluous. Whether a stage takes up any time is immaterial. It is sufficient that they happen at distinct times, which they must if the idea of set formation based on *already existing* sets is to make any sense, and that the before-relation on these instances of times is a well-ordering, which is also part and parcel of the iterative conception. For then there must be a positive amount of time separating the instant of time where a given stage is completed from the next stage, and then the same argument can be run with those intermissions instead.

As hypertasks are impossible, the iteration of set making cannot go beyond the countable stages. So even if we set aside the other problems with temporal Platonism, mentioned above, the iteration could not have the strongly unbounded character which Cantor (1962) and his followers imagine.

An alternative to the temporal interpretation is to accept the dynamic language but take it as a metaphor for a certain kind of dependency relation.[13] The inspiration could be the relation of dependency among theorems of a formal system. A theorem whose proof relies on another theorem must be proved later than that theorem, but on a Platonic understanding of mathematics this does not mean that those propositions achieved theoremhood at different times. It is just that one of them comes later than the other in the order of dependency.[14]

The story would then be that for some sets xx, the set of those sets, $\{xx\}$, depends on xx and owes its existence to them but exists "as soon as" the xx do. This would seem to create the right kind of balance. On the one hand, there is no problem getting beyond the countable stages, as there is no time consuming set creation in play. And on the other hand, we still get the benefit that temporality was supposed to deliver, namely the well-foundedness with which Russell's Paradox can be avoided.

[13] One source is (Potter 2004). However, he does not develop the position much and the presentation here does not follow him.
[14] For critique of this way of justifying the iterative conception see (Incurvati 2012).

However, the balancing act does not succeed. A paradox different from Russell's arises. The problem arises when we ask why, according to this story, there are any sets at all. Why could it not be the case that the sets xx exists but $\{xx\}$ does not? In particular, when xx is the empty plurality, why does the empty set exist? We have been explicit about a necessary condition for $\{xx\}$ to exist but not about a sufficient condition. If that extra needed condition is an act of creation, nothing has been gained.

This answer to the question does seem to be implicit in the idea of dependency. That $\{xx\}$ exists "as soon as" the xx do, presumably means that there is no extra condition; for $\{xx\}$ to exist it takes nothing more than that the xx exist. This leads directly to contradiction: Let xx be all the existing sets. Then $\{xx\}$ should exist. But as no set contains itself, $\{xx\}$ is not among the xx and hence they are not all the sets, contrary to assumption.

There has to be something more that it takes to be a set. The story of dependency does not in itself deliver such a criterion and hence does not give us any reason to believe in a universe of sets that satisfies ZFC, rather than believing that there are no sets at all.

I think it is safe to say that the suggestion of a changing World of Ideas is not a viable way to reconcile Platonism and contemporary mathematics, whether we take it literally or metaphorically. It had to be considered, however, because lack of change was an implicit assumption in the Russellian argument to contradiction and not, on the face of it, a definitorial part of Platonism. By the same token, a second and a third assumption must be considered. The second is the assumption of determinateness, the characteristic that for each Platonic entity and each property (excluding relational properties to objects exterior to the Platonic world such as *is thought of by Peter*) it is timelessly and non-vaguely the case either that the entity has the property or that it does not. That is the ontological underpinning for bivalence and, in particular, that the Russell Set either is or is not an element of itself.

Again we have an assumption that is not strictly contained in the definition of Platonism but nevertheless seems to follow from it. Having postulated the existence of a changeless world to provide the truth makers for all mathematical truths, one is in the worst possible position to deny determinateness and bivalence. When mathematical statements can be taken at face value to be about the objects they purport to be about, it is very difficult to see how they could possibly fail to be either true or false. And, of course, taking that route, one would no longer be in the business of defending classical mathematics, which has bivalence as a core assumption.

The third assumption is one of maximality, namely the assumption that was expressed above by saying that the World of Ideas contains all possible sets. If we hold on to Platonism but reject this auxiliary hypothesis, can paradox be avoided?[15]

A simple-minded attempt is to say that Russell's Set may simply happen to not exist and that the burden of proof lies on Russell, if he should want to show that mathematical Platonism is inconsistent along with Frege's axiom system, to show that it does. That line of thought will not get one off the ground, though. For in the absence of a good *specific* reason for doubting the existence of Russell's Set, one could not know if not also, say, the number 7 (in its set theoretic coding, von Neumann style[16]), would happen to be missing from the Platonic world and thereby the truth maker for sentences referring to that number.

The ontological proposal has to be coupled with a semantic proposal to gain traction. If it might be the case that the Platonic 7 does not actually exist, merely possible 7's must be allowed to play the role of truth makers (in the actual world). It must be admitted that some of the theorems of ZFC are false when interpreted literally and only true when interpreted with implicit boxes and diamonds. An assertion that there exists an x such that ϕ must be translated into the assertion that possibly there exists an x such that ϕ, while the similar universal claim must be interpreted to mean that it is necessary that for all x, ϕ (Linnebo 2010, 155).

In itself this semantic thesis seems quite reasonable. We are looking for something – anything at all – in the ontology we are already committed to, prior to considering mathematics, that can be used as referents for mathematical terms and truth makers for mathematical sentences. Counterfactual possibilities are something we are already committed to if we do not have an extremely counterintuitive idea about what is in Being.[17] (Although, for reasons to follow, I do not agree with Linnebo in general it may be worth foreshadowing that in this particular respect the conclusion of this dissertation agrees with Linnebo's position.)

[15] As far as I know the first proposal in this direction was (Parsons 1975). (However, Parsons is vague about how to interpret the modalities that are at the center of his proposal. So it is *on one interpretation of* Parsons that he is the first to make a proposal in this direction. On another interpretation, he does not make that proposal at all.)

[16] See (von Neumann 1923).

[17] Involving merely possible objects may deliver a solution to Benacerraf's problem – see (Benacerraf 1973) – for according to this thesis we do not need epistemic access to the actually existing sets. In fact it is of no consequence for mathematics which sets happen to have actual existence. The epistemological problem is instead one of knowing what is possible. That may be a simpler problem. At any rate, it is a different problem! I shall not go into it here (although the next section is related to this problem), as I think that the bigger problem is on the ontological side.

That all mathematical objects that could exist, do exist and exist by necessity, is a thought that often goes together with Platonism. This "modalism" implies that that is not the case. Instead, necessarily, there exist some entities that are not collected into a set, but could be. For instance, all the finite ordinals could exist but, "by chance", fail to form a set.

This means that a set does not exist solely by virtue of its elements existing. It is not the case that as soon as God had created the empty universe, so to speak, the empty set existed, along with the singleton of the empty set, etc. It takes something more to be a set. So there is a "part b" of the analysis of "forms a set": a given plurality forms a set iff a) the elements of the plurality exist (or are "available") and b) something else. What could that something else possibly be? It is a trivial observation, and yet it presents a problem for the Platonist, that part b would have to be something contingent; neither necessarily satisfied nor impossible. I can think of nothing else than it being some sort of act of creation. That route brings us back to the problematic temporal interpretation of the iterative conception, and in addition it suggests a creating subject, undermining the abstractness of the mathematical objects.

In an attempt to avoid this conclusion, Linnebo proposes that the modality is not to be taken in a metaphysical sense. All possible sets do, in fact, exist. However, they do not form the set of all sets to which restricted comprehension can be applied so as to imply the existence of the Russell Set, because *individuation* of sets happen in a stage-by-stage process:[18]

> I understand the above modalities in terms of a process of individuating mathematical objects. To individuate a mathematical object is to provide it with clear and determinate identity conditions. This is done in a stepwise manner, where at any stage we can make use of any objects already individuated in our attempts to individuate further objects. In particular, at any stage we can consider a plurality of objects already individuated and use this to individuate the set with precisely these objects as elements. A situation is deemed to be possible relative to one of these stages just in case the situation can be obtained by some legitimate continuation of the process of individuation. (Linnebo 2010, 158)

The first thing to note is that Linnebo seems to appeal to the work of an agent, or rather several agents. They do not create the sets, but it is up to them to make them sufficiently distinct. As a set is constituted by its elements, the agents can only make it clear *which* set a set is when it is already clear which sets its elements are and therefore the individuation must happen in a certain order.

[18] See (Fine 2005) for a related approach.

It can be questioned whether this version of modalism is really Platonistic. A cautious criticism is to say that the letter of the definition is satisfied as we only required the *existence* of mathematical objects to be independent of physical and mental entities, and only the spirit of the definition is not. However, this may well be too cautious. For on Linnebo's account, an explanation of which sets there are and in particular which supposed sets do not exist must involve the agents. The existing sets are those that the agents can potentially individuate. The explanation of the lack of a universal set is that the agents are unable to complete the individuation of all the sets.

That leads me to the more substantial problem of this theory's implausibility. I see no other motivation for the modality of individuation and the role it plays than the desire to avoid paradox. Why is the identity of a set dependent on the actions of agents? If it is assumed to lead an independent life in the Platonic realm it would only be natural, in the absence of good reasons to the contrary, to also assume that its identity properties are a matter internal to that realm.[19]

Linnebo's proposed solution to Russell's paradox plays on a separation of existence and identity: sets exist independently of agents, but agents supply the identities. However, according to Quine's (1969) famous slogan "no entity without identity", existence and identity cannot be separated for *any* kind of entity. And for sets the prospect of doing so is only worse than in the generic case. Sets are constituted by (the identity of) their elements, so to exist as a set just is to have the clear criterion of identity that the set is identical with exactly those sets that have *those* elements. To uphold the proposal, an answer must be given to the questions, how and in what sense a set can exist prior to its elements being individuated, and I do not think one can be given.

In another paper Linnebo (2009) makes a distinction between a semantic and a metaphysical sense of "individuation". To individuate semantically is to determine what is to count as the same object for the purpose of fixing the meaning of a word that is intended to refer to that object. Metaphysical individuation is a matter of what grounds identity facts independently of language.[20] Exactly this distinction spells problems for Linnebo's account: is the process of

[19] From conversation with Linnebo I have learned that he does not intend that acts of agents play a role in his theory. The word "interpretation" is meant to take its meaning from the act of understanding a term, but abstracted from this original meaning to such a degree that no agents are involved. This, however, seems to be to abstract *away* all meaning of the term, and it is not clear that there is a coherent position here. If we abstract away the subjective element of understanding in "interpretation" are we not just left with that which model theory has interpreted "interpretation" as, namely a platonic set, making the justification circular?

[20] At least it can be defined like this by the platonist who believes the objects to exist independently of language.

individuation to be understood in the semantic or in the metaphysical sense? If it is the semantic, it is clear why agents have a role to play, but entirely unclear why that effort has any impact on the platonic sets and hence can be used to avoid paradox. And on the other hand, if it is metaphysical, the platonist cannot justify the appeal to the work of agents, and in particular their inability to individuate the universal set, in his paradox blocking.

Further problems arise when Linnebo considers the question of which sets are actually individuated right *now*. He considers two ways of answering it, this being the second:

> As science progresses, we formulate set theories that characterize larger and larger initial segments of the universe of sets. At any one time, precisely those sets are actual whose existence follows from our strongest, well-established theory. (Linnebo 2010, 159fn21)

This answer is not true to the requirement that for a set to be individuated, all its elements must already have been individuated. For when set theorists "move" from one model of ZFC to a larger one, they take a leap of uncountably many stages at once, and the requirement would only be satisfied if for each stage whose sets have been individuated, there was an instant of time where it was the newest. We are back at the problem that only a countable infinity of stages could have been realized, and even that only on pain of being committed to supertasks. Let us consider Linnebo's first answer:

> The most plausible response to this follow-up question is, I think, that set theorists generally do not regard themselves as located at some particular stage of the process of forming sets but rather take an external view on the entire process. It therefore would be wrong to assign to ourselves any particular stage of the process. (Linnebo 2010, 159)

This is to dodge the question and does not accomplish more than moving the problem back a step. Sure, the set theorists may not themselves be engaged in the process, but rather describe one that some other "individuators" conduct or counterfactually could conduct. However, then essentially the same problematic question re-emerges in another form: who are the individuators and what exactly do they have to do to individuate a set? To this question, Linnebo only leaves us with the above answer *mutatis mutandis*, which is just as unsatisfactory here as it was before.[21]

[21] In conversation, Linnebo has openly admitted to dodging the question and defended doing so. He describes an agent who in detail goes through the first few stages of the hierarchy, then "gets the hang of it" and imagines how the rest of the process goes without filling in

In conclusion, I see no satisfactory way of cashing out the metaphors of the iterative conception and thereby no way of sustaining the method of paradox blocking that is the standard in contemporary mathematics. Only one very *un*satisfactory way is left. It is a version of modalism where certain aspects are simply left unspecified. It is not explained what the sufficient condition for some entities to form a set is, nor what a set *is*, nor what the relation of *being an element* of amounts to. Instead there is just an appeal to the enormity of the space of possibilities along the lines of "yeah, those are some really weird possible worlds, but it is *possible* that there were such a world, right?". The possible worlds contain some objects that we call "sets" and a relation we call "membership", but these words are arbitrary and we can give no story about how the objects and the relation are akin to collections and the relation of one collection being included in another. They, whatever they are, just happen to satisfy the axioms of ZFC.

A possible world in which some "sets" and a "membership" relation exist does not stand in a dependency relation to a possible world with fewer "sets" and a smaller "membership" relation in any sense of "dependency", temporal or otherwise, that anyone has succeeded in articulating. If one nevertheless appeals to such alleged possible worlds, one occupies a position very similar to formalism: the vocabulary of the language of set theory is left uninterpreted, and we could just as well follow Hilbert's suggestion and use "chair" and "glasses of beer" as "set" and "is an element of".[22]

Accepting the iterative hierarchy on that background is to base the mathematical science on the mere negative fact of the absence of a conclusive argument that worlds containing *something* satisfying the axioms of ZFC are not possibility. I think that mathematics, supposed to be the epitome of epistemic certainty, deserves a firmer footing in possibilities that we can at least fathom.[23]

The next two sections are concerned with supporting the claim made here, that it is not even clear what sort of possible worlds classical mathematics is about.

all the specific details and *therefore* is no longer anywhere specific in the hierarchy. I do not find this explanation helpful. This agent does exactly the same as the set theorist: engages in the fantasy of *someone* going through a transfinite process. We, who are looking for an underpinning of this fantasy, gain nothing from being told about one more fantasist.

[22] This, I believe, is the problem with the positive proposal in (Incurvati 2012).

[23] Two different projects can be distinguished: Developing and justifying a mathematics that is secure enough that we dare to claim that theoremhood implies truth for something in Being. And investigating the consequences of more or less arbitrary assumptions without a claim to high epistemic security. My project here is the former, but I do not dispute that classical mathematics is a legitimate way to pursue the latter project.

1.2 Transfinite ordinals

The modalist bases mathematics on the possibility, for each initial segment of the iterative hierarchy, that that initial segment exists. Let P be the proposition that all the countable ordinals and the first uncountable ordinal, ω_1, exist. Then specifically, she is committed to the possibility of P.

The dialectical situation is that I have no convincing argument to the effect that possibly P is false. I am agnostic. But it seems quite reasonable to demand of the classical mathematician that she can produce some positive reasons for believing in possibly P. The epistemic possibility of the metaphysical possibility of P consisting in the absence of proof of the metaphysical impossibility of P is too shaky a foundation for mathematics.

In general, it is my contention that to be warranted in making a positive assertion about the metaphysical possibility of some proposition one must at least be able to describe, or point to, or imagine, or something along those lines, a scenario in which the proposition is true. That is a reasonable demand of a defense lawyer in a trial claiming that it is a possibility that someone different from his client is the murderer, and it is reasonable in the present context. I will try to explain why I think that such a scenario is yet to be provided in the case of P.

Let us set out on the way to Cantor's Paradise. The journey begins with all the finite ordinals:

$$1, 2, 3, \ldots$$

Could they all exist together? I am already skeptical but will assume it until Section 1.4. At least it is pretty clear *what* it is I am asked to believe the possibility of; I can comprehend each element of this infinite set. (The strict finitist may disagree, but he is not my dialectical opponent right now – he will be briefly in Chapter 4.)

If that is possible, it is certainly also possible that one more number, ω_0, exists together with them. Then we are into the transfinite, and having waived reservations about that, it seems undeniable that we must also accept the possibility of a good piece more of the road, such as this:[24]

$$\omega_0, \ldots, \omega_0 \cdot 2, \ldots, \omega_0 \cdot 3, \ldots$$

Again, the classical mathematician must be granted that she succeeds in giving an adequate description of the possibilia that she wants us to acknowledge.

[24] For definitions of ordinal addition, multiplication and exponentiation see, for instance, (Devlin 1993).

Each element of the sequence so far has a name and can be thought of individually, and with a little imagination one can produce a metaphorical image of the entire sequence, such as a road vanishing in the horizon. The same holds further on:

$$\omega_0^2, \ldots, \omega_0^3, \ldots, \omega_0^{\omega_0}, \ldots, \omega_0^{\omega_0^{\omega_0}}, \ldots$$

It actually holds quite a bit further on, for when we run out of possibilities with exponentiation, there are techniques for bringing us further that still allow naming of each ordinal and brings no new philosophical troubles, resulting in the so-called Veblen ordinals (Veblen 1908).

However, no matter how creative we are in devising systems of notation, we have only progressed an infinitesimal part of the distance to the first uncountable ordinal. The reason, of course, is that any system of notation can only name countably many ordinals and by definition there are uncountably many below ω_1. So even if, for any system of ordinals you may describe to me, I accept for the sake of argument, that *they* could all exist, you have still not explained to me, what it is you want me to believe when you ask me to believe P. It's like planning to drive from Paris to Beijing and asking for directions, but only ever getting more and more detailed instructions on how to get out of your driveway, together with a cheerful assurance that after that accomplishment, you should just proceed in the same way and you shall find the Forbidden City.

That was a preliminary way of making my point. It can be made with more precision by taking a closer look at the method for getting larger and larger ordinals. Cantor "created" the ordinals using two "principles of generation". The first states that for every already formed and existing number, a new number can be created by adding a unit to it. The strong principle is the second one, according to which "if any definite succession of defined integers is put forward of which no greatest exists, a new number is created by means of this second principle of generation, which is thought of as the limit of those numbers; that is, it is defined as the next number greater than all of them".[25] The operative word is "created". Presumably, Cantor did not mean it literally and we have already rejected a dynamic understanding of the hierarchy, so neither can we. Instead, we must understand the demand on a succession of numbers to be that those numbers can exist together. So the second principle becomes "For every possibly existing succession of numbers of which none is the greatest, a new number could exist together with them, defined as the next, greater than them all".

[25]"[W]enn irgendeine bestimmte Sukzession definierter ganzer realen Zahlen vorligt, von denen keine größte existiert, auf Grund dieses zweiten Erzeugungsprinzips eine neue Zahl geschaffen wird, welche als *Grenze* jener Zahlen gedacht, d.h. als die ihnen allen nächst größere Zahl definiert wird." (Cantor 1883, §11, original emphasis).

On that interpretation I have no objection to the two principles. Between them, they just say that if some numbers can exists together, then those numbers plus one more could too. Only a strict finitist would deny that. The question is what could exist together.

When we have a system of notation for a set of ordinals, it is relatively clear what believing in their possible co-existence amounts to. Hence, the application of the second principle to such sets is not as objectionable as the more liberal applications that Cantor allows. However, as noted, it does not get us out of the domain of the countable.

The more liberal and daring way of using the second principle is when a succession is considered "formed" simply because it, allegedly, consists of exactly those ordinals that satisfy a given predicate. The only way to get ω_1 is to take the predicate of being a countable ordinal, turn it into a set by comprehension and apply the second principle to that. Cantor gives us no reason to believe that such a set could exist, he merely presupposes it. It is of course not legitimate to use the second principle like this on any predicate. If the predicate applies to all possible ordinals, as for instance the vacuous predicate of being self-identical does, we run into Burali-Forti's Paradox (Burali-Forti 1897): the new ordinal that the second principle produces must be larger than all ordinals including itself. How do we know that "countable" is not such a predicate? We do not. Only if there is a possible uncountable ordinal, is it the case that "countable" does not apply to all the possible ordinals that "self-identical" applies to. So the possible existence of such an ordinal is only guaranteed by Cantor's method in a viciously circular way.

Cantor's attempt at convincing us of the possible existence of ω_1 in effect takes the following form. Let ϕ be a predicate such that we do not know whether all possible ordinals satisfy ϕ. Consider the set of all ordinals that satisfy ϕ and apply the second principle to it. Call the resulting ordinal α. By *reductio* it follows that α does not satisfy ϕ. Ergo, ϕ does not apply to all possible ordinals.

It is conviction created *ex nihilo*. From the mere absence of a *known* contradiction in the assumption of all ordinals satisfying ϕ existing together, Cantor wants us to conclude that they can.

I think we should remain skeptical about the possible truth of P because we have but the flimsiest of ideas of what the truth of P would amount to. No intuitive understanding can be provided, for any attempt at doing so involves transfer of spatial and temporal metaphors to a domain where they clearly break down. The grasp that we have of P is purely conceptual and purely negative; it is based on nothing but the negation of the property of being countable. The fact that no contradiction has, as yet, been deduced from the linguistic description of P, is to my mind not sufficient assurance that I would base mathematics on it.

I am also skeptical about the very first use of Cantor's second principle, the one that takes us into the actual infinite. I will return to that below, after a section where the focus is shifted from transfinite ordinals to transfinite cardinals.

1.3 The continuum hypothesis

According to the believer in the possibility of actual infinity, there are two routes into the transfinite cardinals. One goes via the ordinals: First we take the predicate of having finitely many predecessors, negate it, and turn it into an ordinal by Cantor's second principle. That gives us the first ordinal that has infinitely many predecessors, ω_0, and as a consequence (by abstracting from the order of the predecessors of ω_0) the first infinite cardinal, $\aleph_0$. Then, similarly, we take the predicate of having at most $\aleph_0$ predecessors, negate it, and turn it into the first ordinal to have more than $\aleph_0$ predecessors, ω_1, with its associated cardinal $\aleph_1$. And so it continues, $\aleph_2, \aleph_3, \ldots, \aleph_\omega, \aleph_{\omega+1}, \ldots$, to name the first "few".

The second route is by the notion of powerset. The procedure of going through the elements of a collection, selecting or deselecting each one in order to form a subset, is idealized to infinite sets, and it is further assumed that there can be a set of all the possible results of this infinitary "procedure". As Cantor's Theorem (1891) tells us that the powerset of any set is larger than that set, repeatedly taking powersets is an alternative engine of transfinite cardinal production: the set of natural numbers gives us one cardinal, its powerset another cardinal, the powerset of the powerset of the set of natural numbers a third cardinal, etc.

If one believes that we are dealing with a clear conception of the infinite and in particular that we are in possesion of a definite idea about what the world would be like if P were true, then one must believe that these two routes lead to the same place. Any cardinal "produced" by the ordinal method must be comparable to any cardinal "produced" by the powerset method, i.e. it must be the case that either the first is smaller than the second, they are identical, or the first is larger than the second. In particular, there must be an answer to the question whether the cardinal number of the powerset of the set of natural numbers is identical to $\aleph_1$.

However, the answer to this question, the question of the truth value of the continuum hypothesis, is yet to be found. I am not only referring to the fact that the continuum hypothesis has been shown independent of the axioms of ZFC.[26] Rather, I mean it in the broader sense that, as far as known, no answer to the question can be informally deduced from the intuitive ideas of Cantor's generative principles and powersets. Or in more familiar terms: no suggestion for adding an axiom to ZFC which would settle the question and which is intuitively true, assuming the mentioned ideas, has won widespread support.[27]

As far as is known, Cantor's generative approach and the power set approach to the transfinite do not match up. I will let Quine enter the stage again, just to repeat his one line: "no entity without identity". The theory of the transfinite is in a condition where it is not at all clear *what* it is we are asked to believe in the possibility of. The continuum hypothesis is not just an interesting open research question; its undecidability is a philosophical problem for those who believe in the possibility of infinity. Of course, it may be the case that the question is not undecidable, but just undecided. If that changes in the future, the believer in the possibility of actual infinity will have a somewhat stronger case. However, in the present state of affairs the openness of the question is another reason for being skeptical.

1.4 Countable infinity

Above I have given reasons for being skeptical about actual infinity beyond the countable. Now I will aim the guns at this, the smallest kinds of actual infinity, itself and try to convey an understanding of why there is room for rational skepticism about its possibility, even though no explicit contradiction is known to follow from the assumption of its existence.

Let me make it clear that I readily concede the epistemic possibility of the metaphysical possibility of actual infinity, for I do not have a proof that it is not metaphysically possible. I also do not think that actual infinity is unintelligible. In one sense it is perfectly clear *what it is* they claim, those who argue for the metaphysical possibility of actual infinity. I understand what it is for the world to be finite, and therefore I understand the negation of that proposition. I understand it, but only conceptually. I have no intuitive understanding of the possibility of actual infinitude. (What about the road vanishing in the horizon? That is merely a *metaphorical* image of actual infinity, for it is not an actual image of actual infinity. To get the sense of infinity, the still image

[26] Gödel (1940) showed that the negation of the continuum hypothesis does not follow from ZFC, and Cohen (1963, 1964) proved that neither does the continuum hypothesis itself (both assuming that ZFC is consistent).

[27] For detailed discussion of the present situation, see (Woodin 2001a), (Woodin 2001b) and (Koellner 2013).

is not enough; I have to play a movie in my head of me driving along the road, never meeting an end. That is a movie illustrating potential infinity.)

Let me take Benardete as my dialectical opponent at this point in the discussion. He writes the following:

> I am prepared to agree with the finitist rejection of aleph-null as a cardinal *only on the assumption that it is absurd to speak of aleph-null apples*. But that assumption is a mistake. The finitist is obliged to say that he knows *a priori*, apart from any empirical or scientific evidence, that there exist only a finite number of stars in the heavens. If it is unintelligible and meaningless to speak of an infinite number of stars, then it follows that we know *a priori* that a spaceship which is launched to explore the heavens must be at some precise finite time in the future (travelling at a uniform velocity) eventually encounter each and every star that now exists. Finitism is thus seen to be a form of apriorism at its worst. (Benardete 1964, 30–31, original emphasis)

I reject the claim that I have committed myself to such an obligation. Benardete slides from something being “absurd” to it being “unintelligible and meaningless”. The former relates, presumably, to metaphysics, the latter to semantics. I think it is absurd to think that an infinity of stars could exist (as I shall explain in due course), but I do not think it is unintelligible and meaningless to suppose so. A totally absurd scenario can intelligibly be described. He also slides from an infinity of stars being metaphysically impossible, which is what the finitist is committed to,[28] to that we know (present tense) *a priori* that the number of stars is finite. I think that it is metaphysically impossible for an infinity of stars to exist, but I do not claim to *know* so.

Let us help ourselves to an inter-possible-world-travelling spaceship and go visit one of these worlds where there are infinitely many stars. In the world we arrive at there are, unsurprisingly, infinitely many inhabited planets orbiting a subset of the infinitely many stars. Landing on one of these, the planet of Zvyagel, to explore further, we immediately notice the immense wealth of the Zvyagelians. They all live in luxurious houses that are best described as castles, they have more delicious food than they could possibly eat and whenever a yacht of theirs gets wet they discard it and switch to a larger one. First assuming that the Zvyagelians are in possession of superior production technology, we are corrected when we, after having been guests on the planet for a while, learn how to communicate with the Zvyagelians. They tell us

[28] Benardete's “finitist” should therefore not be confused with the “strict finitist” mentioned elsewhere in this dissertation. The former believes that actual infinity is impossible, the latter rejects even potential infinity.

that until a century ago they suffered from the same problems of insufficient resources as we do, and that the solution came not from an engineer but from an economist. She came up with the scheme, which was quickly agreed on and executed, that they should contact a hundred other inhabited planets and ask them to send everything they produced. They all obliged and have since then each sent a rocket every month containing their entire production of food and other goods, allowing the Zvyagelians to live a life of abundance. The marvelous thing is that no threat of military power was sent along with the request. The inhabitants of the supplying planets all acknowledged, free of any duress, that the scheme was beneficial to all involved parties. For along with the request came the instruction to pass on the same request and instruction to a hundred other planets, so that each supplying planet was itself supplied by other planets who in turn was supplied by still other planets and so on.

This scenario of a well functioning pyramid scheme I find absurd, and even though it is neither unintelligible nor meaningless nor involves any contradiction that I am aware of, I doubt very much that it is possible. I find it too absurd and doubt it too much to be comfortable having the queen of the sciences be associated with such tales of fantasy.

Coincidently, during our visit to Zvyagel, the Zvyagelians receive the great honour of hosting the annual meeting of the Society of Gods which has denumerably infinitely many members.[29] In addition to their usual business, the gods have decided to use this year's meeting to hold a fair infinite lottery.[30] The Zvyagelians can enter the lottery by buying a ticket with any natural number printed on it. The gods therefore produce an infinity of balls, one for each of the natural numbers and each marked with that number, to use for drawing a winner. All the balls are placed in an urn. That this is a regular urn of finite size poses no problem, of course, for the n'th ball has a volume of $\frac{1}{2^n}$ cubic centimeters.

The membership cards of the Society are also numbered by the naturals, and according to that numbering the urn is passed from god to god in reverse order so that in the span of one minute the urn has been in the possession of each member of the Society. God number 1 will be passed the urn at $t = \frac{1}{2}$ and it is his job to remove all but one ball from the urn before $t = 1$; the remaining ball contains the winning number. Before that, the urn will have been at god number 2, who gets it at $t = \frac{1}{4}$ and reduces the number of balls to 2. The third god will get the urn at $t = \frac{1}{8}$ and he must leave 4 balls. In general, god number

[29] The minutes of a previous meeting of the Society are to be found on page 259-260 of (Benardete 1964).

[30] They are inspired by (de Finetti 1974) where the idea is discussed in the abstract without any indication of how to actually carry out such a lottery.

n is handed the urn at $t = \frac{1}{2^n}$ and has $\frac{1}{2^n}$ minutes to reduce the number of balls in it to 2^{n-1} and, if $n \neq 1$, pass it on to god number $n - 1$.

The gods do not simply give themselves the instruction *when god n receives 2^n balls, he must leave 2^{n-1} random balls in the urn.* If they did, they would not have ensured that each god n actually receives 2^n balls. They might all receive the urn still containing all the balls, thus being unable to follow the instruction, because all the previous gods have received the urn still containing all the balls and been unable to follow the instruction. The gods guard themselves against this danger by instead giving themselves this more complex instruction: *if god n receives 2^n balls, he must leave 2^{n-1} random balls in the urn. If he receives more, he must leave the lowest-numbered 2^{n-1} balls in the urn.* If any god were to follow the consequent of the second part of the instruction because he receives too many balls, he would remove balls in a non-random way, thus ruining the experiment. However, the complex instruction has the effect that this does not happen. The instruction ensures that each god leaves the right number of balls in the urn, no matter what. Ergo, each god will act on the first part of the instruction, and the second part is never actually used.[31] That way the gods ensure that the experiment runs as intended.

Each god, being a god, has no problem executing his task. All he must do is to choose some given finite number of balls to leave in the urn and he has a positive amount of time to do so.

Indeed, they all succeed. After the one minute has passed, the urn contains but one ball, chosen completely at random. One of the gods takes the last ball out of the urn and reads the number, k_1, out loud to his colleagues.

They are astonished! What a surprisingly low number! There are infinitely many larger numbers that could have been the winner instead and only $k_1 - 1$ lower numbers, and none was more likely than any other!

They all swear they have made their selections without bias. Unlike mere mortals, they have the ability to completely ignore part of their knowledge at will when making a decision and they all used that ability to disregard the sizes of the balls.

Assuming that such an unlikely result must be a one-time fluke, the gods decide to repeat the experiment. They run it a hundred times. Worried that the unequal size of the balls might somehow have influenced the first result in spite of the gods' special doxastic ability, they vary the particulars of how the draws are conducted. A few examples: In the 17th experiment they write the numbers on index cards and spread them out, face-down, over an infinite area.

[31] In this respect they take inspiration from (Laraudogoitia 2011).

In the 52nd experiment they opt for using inflatable balloons; they start out having different sizes just like the balls, but as soon as there is only a finite number left they all change to having the same size; namely, at any given time, one cubic centimetre divided by the number of remaining balloons. In the 90th experiment the gods do not directly choose which numbers to remove and which to leave; instead they divide the remaining numbers into two piles and flip a coin to decide which pile is removed.

Nonetheless, the same thing happens again and again. The results are different in that the winning numbers, $k_2, k_3, \ldots, k_{100}$, are different, but alike in the respect that those numbers are all shockingly low.

The gods end their meeting in a state of bewilderment. The lotteries have been conducted in a completely fair way, with no god having a bias towards low numbers, and yet the results were unbelievably low every time.

This continuation of the story about Zvyagel has the same purpose as the first, namely to show that the assumption of the possibility of actual infinity leads to absurdities that are almost as bad as contradictions. Here the absurdity is that of a necessary surprise; any result is much lower than one would expect. That, anyway, is the somewhat naïve way that I have portrayed the gods as thinking. In the rest of this section we will consider attempts to bring out what is remarkable about the lottery in more precise ways.

The supertasks literature consists almost exclusively of papers that draw consequences from the assumption that supertasks are possible. These include the possibility of stopping an object by intentions alone and prior to all of the intended actions (Benardete 1964); spontaneous self-excitation of a static system (Laraudogoitia 1996); creation *ex nihilo* (Laraudogoitia 1998); forcing a spaceship to arrive from infinity (Laraudogoitia 2011); adding and removing balls from an urn in such a way that there are constantly added more than are removed but resulting in the urn being emptied (Allis and Koetsier 1991); and that player A has a winning strategy in a game where all player B has to do to win is to repeat whatever A says, whenever A utters "zero" or "one" (Bacon 2011). In all cases, the consequences are accepted as merely surprising and exotic and never used for a modus tollens. I think the latter is a move worth considering.

It is a very difficult case to make, for in a debate concerning whether something is metaphysically possible or impossible, the burden of proof will often be assumed to rest with the advocate of impossibility, and if his opponent in this case decides not to accept anything but an explicit contradiction as a consequence worth taking note of, and instead bites the bullet on all "surprising consequences", there is little he can do to move her. That is one reason I will still not claim to reach an unequivocal conclusion, but only to cast doubt on the possibility of actual infinity. In addition, as we will see below, explaining why the *surprising* consequences should rather be seen as *absurd* consequences depends on some intuitions that can be difficult to articulate precisely, and those difficulties can be exploited by an uncooperative interlocutor. I believe that the infinite lottery makes it ever so slightly easier for the supertask skeptic to make his case than with previously considered supertasks.

Let us settle on some names for the two combatants: the believer in the metaphysical possibility of reverse supertasks will simply be called "Believer" and her opponent, with whom my sympathies lie, "Disbeliever". Disbeliever thinks that describing the necessary surprise as "remarkable" is an understatement. Believer thinks that it is an overstatement and that the gods should not be surprised at all.

One tentative way of trying to bring out what is remarkable about the phenomenon is to say that no rational expectation can be formed about the outcome of the experiment. The experiment is a stochastic process for which no probability distribution exists. Therefore, it is peculiar that the experiment has outcomes, for by repeating the experiment one obtains an empirical distribution function that would seem to offer a basis for rational expectations about future instances of the lottery.[32] That gives rise to a dilemma, namely that neither an affirmative nor a negative answer to this question seems to be acceptable: after having done the experiment 99 times, should the gods expect k_{100} to be lower than $\max\{k_1, \ldots, k_{99}\}$? A negative answer seems unacceptable because that would be to disregard what seems to be relevant empirical evidence. And a positive answer seems unacceptable because that would amount to assigning larger probabilities to some outcomes (those lower than the max-

[32]There is no uniform probability distribution on the natural numbers according to Kolmogorov's (1933) original definition. Weakened notions of "probability distribution" have been proposed, see (de Finetti 1974) and (Wenmackers and Horsten 2013). However, the probability distributions for this lottery, according to those definitions, only assign non-zero (or non-infinitesmal) probabilities to certain infinite sets. So they do not assign probabilities that can give an indication of the size (or "order of magnitude") of the outcome to be expected. Hence, it is peculiar that it has outcomes *with sizes*, for by repeating the experiment one obtains an empirical distribution function that can form the basis for rational expectations about *the size of* future outcomes of the lottery. Therefore, these alternative definitions do not solve the problem. Rather, accepting one of them makes it possible to express the curious nature of the infinite lottery using a precisely defined notion of probability: for any actual outcome k, the probability that the outcome would have been higher is 1 (de Finetti) or 1 minus an infinitesimal (Wenmackers and Horsten). (It could however *not* – in spite of these definitions making a well-defined notion of "expected value" available – be expressed by saying that it is unreasonable that all the outcomes are finite when the expected value is infinite. For that is also the case in the St. Petersburg Gamble (Bernoulli 1954), where it is reasonable.) And, of course, it would also be unreasonable to weaken the definition further so as to allow uniform probability distributions that assign (non-infinitesmal) positive probabilities to finite sets.

imum) than to other outcomes in an experiment where no outcome should be more likely than any other.

Disbeliever sees in this dilemma an absurdity that indicates that the lottery is impossible. Believer counters that a sample of 99 can easily be misleading – even in a finite lottery. An infinite sample would be needed, and with probability 1 such a sample would have no maximum, dissolving the dilemma. Disbeliever interjects that in this case the sample of 99 would not only be *necessarily* misleading, it would be necessarily misleading *in a fixed direction* (the maximum of the sample will necessarily be smaller than the "median of the probability distribution"[33]), and that is absurd. Believer points out that there is no contradiction, refers to the general fact that there are many surprising results in probability theory, and remains unmoved.

As mentioned, the existing literature on supertasks reveals that the possibility of supertasks implies the possibility of a range of rather surprising and strange phenomena. However, in all these cases, the surprising consequences notwithstanding, it is possible to give adequate mathematical descriptions of the scenarios. What is new about the infinite lottery is that it is a supertask that we do not even know how to handle mathematically. Believer may think that the infinite lottery should be impossible *because* no uniform probability distribution on $\mathbb{N}$ exists while other forms of supertasks are possible. Disbeliever would object that the story does not *presuppose* the existence of a uniform probability distribution, but, rather, that the possibility of this scenario only presupposes that (reverse) supertasks are possible; if they are, it seems to *follow* from this possibility that there should be a uniform probability distribution on $\mathbb{N}$. Since there is not, Disbeliever concludes, this is a story about an event that defies description by probability theory, which is an absurdity.

Another way to bring out what is remarkable, which may be an improvement on the first, is to imagine what would happen if the experiment was repeated a hundred times and someone after each experiment, based on the evidence gathered so far, tried to ballpark the outcome of the next experiment by mentioning a natural number such that the outcome is predicted to be lower than that number, say by taking the maximum of the previous outcomes and multiplying it with a googolplex. If this person would tend to be successful, we have the problem already described: some outcomes are more likely than others. And if he tends to be wrong, i.e. if the outcomes tend to increase, that would also be a problem. For it would be very difficult to explain why a permutation of the order of those outcomes should not have been equally likely. The

[33]This is *not* the case in the St. Petersburg Gamble.

experiments would not be independent. Therein Disbeliever sees an absurdity only solvable by rejecting the possibility of (reverse) supertasks. However, Believer has options. She can embrace one of the horns of the dilemma[34] and, to the frustration of Disbeliever, declare that horn for an important and surprising *insight*: “If the gods run a series of experiments, some force will make lower numbers more likely than higher numbers” or “If the gods run a series of experiments, some force will make them dependent”.

A third way is through the concept of “bias”. When considering the gods’ own conclusion that there had been no bias, it seems necessary to make a distinction between intentional and extensional senses of “bias”. On the one hand, the gods had no intention to be biased, nor – we can stipulate (at least for the 17th, 52nd and 90th experiment) – the knowledge required to make intentionally biased choices. On the other, they ended up actually removing larger numbers and leaving a “surprisingly low” one to be the winning number. To imitate a phrase from (Yablo 2000),[35] the gods are “forced by mathematics” to make extensionally biased choices. According to Disbeliever, the first horn of the dilemma implies a commitment to such a force, and by spelling out what that commitment amounts to, he hopes to put more pressure on Believer: Assuming independence of the experiments, the maximum of the outcomes of a series of executions is a guide to the order of magnitude of future executions, which means that the numbers below that maximum have positive probability. *Prima facie* one would think that each ball has a 0.5 chance of being left in the urn by each god, but the positive probability of some outcomes implies that this cannot be the case. The probability that a given number is the winning number is the product of its probability of being left in the urn by each god. For that product to be positive, the probability of being left in the urn by god n must tend to 1 as n tends to infinity. In other words, all but finitely many gods are virtually forced to leave certain numbers in the urn.

Disbeliever also tries to put some meat on the idea of bias in another way. Let $\overline{k}$ be the empirical mean of the outcomes of a number of executions of the infinite lottery. Furthermore, let N be a natural number such that $\frac{2^N}{2} > \overline{k}$. Consider a finite lottery between the numbers $1, \ldots, 2^N$ which is conducted similarly to “the last part” of the infinite lottery, involving N gods: god number N reduces the number of balls to 2^{N-1} in a random way, god number $N-1$ cuts it down to 2^{N-2} and so on. The expected value of the outcome of this finite lottery is $\frac{2^N}{2}$. However, on the one hand, this value is larger than $\overline{k}$. On the other hand, the content of the urn when there are 2^N balls left in an execution of

[34] She cannot say that a finite sequence with some given tendency might be misleading, for that just takes us from one horn of the dilemma to the other.
[35] “Logic stops them.”

the infinite lottery – let U be the set of those numbers – must dominate the content of the urn at the beginning of the finite lottery in the following sense: the function $f : \{1, \ldots, 2^N\} \to U$ that takes each n in the domain to the n'th lowest element of U is such that for all n, $f(n) \geq n$. What this means is that not even the final N gods are able to conduct their part of the experiment in the unbiased way that it should otherwise be clear that they are. The hidden force is able to exert its mysterious influence well into the domain of the finite, where mundane laws of probability should reign supreme. Believer (if she accepted the first horn) may be a little surprised by the extent of her commitments, but can insist that she can take them on, once again noting that no actual inconsistency has been deduced.

Believer also takes issue with the claim that an outcome of the experiment can be "surprisingly low". If an outcome is surprisingly low simply because there are finitely many lower numbers and infinitely many higher, then all numbers are surprisingly low and then none really are. The surprise is naïve because for any possible value of k_1, there are finitely many possible outcomes smaller than k_1 and infinitely many larger, so being surprised that that is the case for the actual value of k_1 is to be surprised by a(n epistemically obvious) tautology. Disbeliever acknowledges that, of course, being surprised by the proposition "there are finitely many possible outcomes smaller than k_1 and infinitely many larger" coming out true, if the experiment was actually executed, would be naïve. But he answers with a distinction between a first and a second order. It makes sense, he maintains, to be first-order surprised about the specific outcome while at the same time being aware that the outcome must necessarily be such as to be first-order surprising and thus not be surprised by being surprised. For Disbeliever, the description of the outcome as a necessary surprise is not self-undermining but rather the simplest and most elegant way to express what is absurd about the lottery. Believer responds that even if such a distinction is reasonable, your first-order surprise must be relative to a subset of $\mathbb{N}$ specified in advance. If you had thought of $\{k_1\}$ in advance, you would be justified in being surprised. And, more generally, if you had specified any finite subset of $\mathbb{N}$ and it turned out to contain k_1, you would be justified in being surprised.[36] But you should not be surprised by *any* singleton. Disbeliever responds with a rhetorical question: if the lottery was executed, would you really not be surprised if the outcome was 7?

Disbeliever has mixed feelings at this point. On the one hand, he feels beaten, out of rational arguments, back at the naïve-seeming reaction of the gods, left with a rhetorical question to which he can add strength in no better way than

[36] Having read (McCall and Armstrong 1989), Believer is cautious not to go on to set up a precise criterion for when surprise is warranted.

by uttering it forcefully. On the other hand, he thinks he is actually back at his strongest and most persuasive point: it would be baffling if there were only 6 balls with lower numbers on them, when there are infinitely many with larger numbers. There just seems to be no way to explain *why* to someone who does not feel the weight of the rhetorical question.

Of course, the point generalizes, as 7 can be replaced with any other number. So Disbeliever concludes that each drawing implies the absurdity of a necessary surprise. And he uses this to dismiss the idea the Believer can take refuge in embracing the second horn of the dilemma. For in addition to dependency implying the existence of an inexplicable causal force between the individual executions of the experiment, the real problem is with the very first execution of the experiment, so the relationship between several executions, such as dependency or lack thereof between them, is, at the end of the day, beside the point and served merely to highlight the problems.

I think this discussion between the two parties comes down to an intuition about whether 7 would be a surprising outcome. And here I side with Disbeliever. Believer can pretend not to understand what it is Disbeliever is getting at. She can do that because it is difficult for Disbeliever to articulate precisely why he has the intuition. However, the fact that Disbeliever's points are largely expressed in pre-theoretic language might be more of an indication that theoretic language at this point in history is not adequate to the task, rather than that he does not have a reasonable point.

Let us take a step back. My overall goal is to cast doubt on the possibility of actual infinity. I have implicitly assumed that if it is possible, then supertasks are as well, and since, arguably, that leads to absurd consequences, there is a *reductio* that should be blamed on the possibility of actual infinity. One might try to block the *reductio* after the admission of the possibility of actual infinity by claiming that the possibility of supertasks is an extra assumption or by pointing out that the story also relies on other controversial assumptions, such as the spectacular abilities of the gods. However, this can be avoided. Just imagine that Zvyagel and all the other planets of the universe are numbered. (This could be the case in a very concrete way, namely if for each planet there existed a capsule somewhere in the universe containing the number on a piece of paper and pointing at the corresponding planet. That way, the planets do not have to be of a particular size to contain the number even if the numeral takes up a lot of space. Only the capsules have to be of unbounded size, and since they can be placed anywhere in an infinitely large universe, that is no problem.) Then it would be a necessary surprise how low Zvyagel's number was.

Then one can try to block the *reductio* by claiming that an actual infinity of abstract objects is possible but not an actual infinity of concrete objects. That is a more interesting objection, as it brings out a premise I am relying on. I have the Aristotelian view that abstract objects must be somehow grounded in concrete objects. Again, I do not claim to *know* that there could not be abstract objects that are independent of concrete objects, but I am skeptical and therefore want mathematics not to rely on such assumptions.

With this, I shall rest my case about only trusting in the possibility of finite worlds. I cannot prove that the infinite is impossible, but I do think that the problems I have pointed out should give us pause. Contra Benardete, I do not think that our current ignorance about the consequences of actual infinity should be taken as warrant for believing it to be possible and available as a basis for mathematics. We should make do with potential infinity. This means, cf. the introduction, that the ultimate accomplishment of justifying classical mathematics is not within our grasp. We must settle for less.

1.5 Intuitionism

We seem to be pushed towards the intuitionism of L.E.J. Brouwer. He aimed exactly for a philosophy of mathematics that satisfies our overall desideratum, not to postulate entities in Being that we would not otherwise believe in, and, in particular, avoids actual infinity. He does so by identifying the subject matter of mathematics with the potential infinity of mental constructions of a creating subject. Brouwer's intutionism will be the subject of our investigation from here, through Chapters 2 and 3, and halfway into Chapter 4. This section will provide an uncritical introduction to Brouwer's thinking.[37] The following two and a half chapters will critically examine two key aspects of intuitionism, namely the theory of free choice sequences and the failure of *tertium non datur*.

Inspired by Kant (1781), Brouwer ontologically locates mathematics in the human intuition of time.[38] The basic building block of mathematical constructions is the so-called empty two-ity, which is the result of fixing on a moment of time, noticing it giving way to another moment of time, and abstracting away the contingent and specific elements of the experience that the subject happens to have at that moment. The construction of the empty two-ity gives us the numbers 1 and 2. That can be iterated by dividing the *now* of the initial two-ity's *past-now* distinction into a "new past" and a "new now" moments, resulting in an object *old past-(new past-new now)* that can play the role of the number three, and so on. According to Brouwer, the mathematical

[37]See (van Atten 2004) and (van Stigt 1990) for more thorough introductions.
[38]See page 8 and chapter 2 of (Brouwer 1907).

universe is limited to what can be constructed in this way. (That is somewhat vague, but that is because Brouwer himself is vague.[39] However, some of the details will be filled out below.)

Referring back to the examples presented at the beginning of this chapter, the sentence “7 is larger than 6” can be interpreted as saying that it takes more construction steps to make the number 7 than to make the number 6. Both Mill and Brouwer, therefore, find references for the terms of that sentence and a truth maker at no ontological price (for someone who already believes in the existence of physical objects and mental constructions).

However, Mill faced the problem that he could not assign to “2 + 2 = 4” the necessity that we are disposed to think that the truth of this sentence has. Brouwer arguably does better on this account. His explication of “2 + 2 = 4” would be as follows: I have constructed a two-ity, then another two-ity and then a four-ity, and succeeded in constructing a bijection between the disjoint union of the two former and the latter. The account of the meaning of the sentence in terms of the behavior of pebbles had the weakness that the laws of nature might change tomorrow, making “2 + 2 = 4” false. A similar weakness may be perceived in the fact that, when I try to construct the truth maker for “2+2 = 4”, I may make a mistake and not get the bijection. However, there is a crucial difference between the physical objects semantics and the mental semantics, which means that the undermining of mathematical necessity caused by the possibility of changed laws of nature does not transfer to mental constructions. A mental construction comes with an *intention* to execute the construction in a certain way. This intentionality implies that there is a normative aspect to constructions, which allows us to say that any *correct* construction of the sum of two and two would necessarily result in four. (I will not consider possible objections here, but see Section 4.3 for a discussion of Wittgenstein’s rule following skepticism.)

Finally, we considered the example of the number 10^{100} above. Mill’s empiricism results in a strictly finitistic mathematics, and is therefore also in that respect inferior to Brouwer’s intuitionism. Even though the creating subject can only ever have completed a finite number of constructions *using* the intuition of time, *reflection* on the intuition of time shows the subject that the future is in principle (in some sense of “in principle”) open-ended and that the series of natural numbers could therefore in principle be extended indefinitely. It is therefore clear that 10^{100} is potentially constructable. Thus, Brouwer’s mentalism provides support for a mathematics of potential infinity but implies a rejection of actual infinity.

[39] See (Kuiper 2004) for an attempt at filling out some of the details omitted by Brouwer.

According to Brouwer, the mentalistic ontology also necessitates a rejection of classical logic.[40] An elegant illustration can be given with the classical proof that there exist irrational numbers a and b such that a^b is rational. It is a proof by cases: Either $\sqrt{2}^{\sqrt{2}}$ is rational or irrational. If it is rational, let both a and b be equal to the irrational number $\sqrt{2}$, and then a^b is rational. If it is irrational, let a be equal to $\sqrt{2}^{\sqrt{2}}$ and let again b be equal to $\sqrt{2}$, in which case we have

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2,$$

i.e. again a rational number. This proof is non-constructive in that it does not inform us which irrational number a has the sought after property. And for Brouwer that is an epistemic point with ontological implications: if we have not constructed an irrational number a and constructed its having the property of being equal to a rational number when raised to the power of an irrational number b, then there is no such number, for there is nowhere else in all of Being to locate it than in our constructions.

The culprit in the classical proof is the very first step, the assumption that $\sqrt{2}^{\sqrt{2}}$ is either rational or irrational in the absence of a construction to support one of the disjuncts. Thus *tertium non datur* is not in general a valid principle.

For Brouwer, logic does not have the central position in mathematics that it has according to the classical mathematician. Logical laws are merely highly general descriptions of the interrelations of constructions. Actually, they are merely highly general descriptions of the *language* that can, imperfectly, be used to convey an essentially language-less construction from one subject to another. An inference rule being valid means that whenever constructions corresponding to the premises are at hand, a construction corresponding to the conclusion can be effected. (This subject will be elaborated on in Chapter 3.)

The non-standard ontology in general and the revision of logic in particular mean that a long range of important classical theorems fail intuitionistically. Another example (that has the virtue of leading us to a few other aspects of intuitionism that need to be introduced) is the theorem that every real number is positive or non-positive, $\forall x \in \mathbb{R}(x > 0 \vee x \leq 0)$. Brouwer gives examples of real numbers for which we cannot assert that it is one or the other.

A prerequisite for these examples is the intuitionistic notion of real numbers. With the exception of the strict finitist, all parties to the debate agree that a real number is an infinitary object. Either it is an ordered pair of actually infinite sets of rational numbers (Dedekind 1872), an actually infinite equivalence class of actually infinite, converging sequence of rational numbers (Cauchy

[40] See (Brouwer 1908).

1821; Heine 1872), or, if you ask Brouwer, a potentially infinite, converging sequence of rational numbers. A real number is the process of a creating subject constructing more and more terms of a so-called free choice sequence (see Chapter 2 for extensive discussion). The terms can be freely chosen by the subject, or he can decide to follow a rule when choosing terms. In the latter case, it must be possible to calculate each term in a finite amount of time for which an upper bound is known in advance. The specific details of the definition of "real number" can be filled out in several different, intuitionistically acceptable ways. In the interest of uniformity throughout the dissertation and for simplicity, we will do this by stipulating that a real number is a free choice sequence $\langle q_1, q_2, q_3, \ldots \rangle$ of rational numbers, such that $|q_m - q_n| \leq m^{-1} + n^{-1}$ for all natural numbers m and n.

A real number for which it can neither be asserted (at present) that it is positive nor that it is non-positive is constructed using a so-called fleeing property, defined by Brouwer (1955, 114, original emphasis) as follows:

> A property f having a sense for natural numbers is called a *fleeing property* if it satisfies the following three requirements:
>
> (i) For each natural number n, it can be decided whether or not n possesses the property f;
>
> (ii) no way is known to calculate a natural number possessing f;
>
> (iii) the assumption that at least one natural number possesses f, is not known to be contradictory.

An example of a fleeing property P is, for a given finite sequence of digits not yet found in the decimal expansion of π and not yet proved not to occur in it, that that sequence occurs beginning at the n'th decimal. Then let the real number a be defined as the free choice sequence the begins with the terms $-1/2, 1/4, -1/8, \ldots, (-1/2)^n, \ldots$, and continues like that as long as no n has had the property P, and stays constant at $(-1/2)^n$ from the first n that has the property P onwards (if there is a such). For the purpose of evaluating inequality statements, we can identify a with the limit of the free choice sequence. (A precise definition will be given in Chapter 2.) Then at any given point in the construction where the choice sequence is still "oscillating", the creating subject is not in possession of a truth maker for either of the sentences $a > 0$ and $a \leq 0$.

This invalidity of a classical theorem leads to the validity of a non-classical theorem, namely that all functions from $\mathbb{R}$ to $\mathbb{R}$ are continuous.[41] Let me illustrate by explaining why this is an illegitimate definition of such a function:

[41] See (Brouwer 1924a).

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

The function would have to map a to a free choice sequence $f(a)$. The first two terms of $f(a)$ could both be equal to $\frac{1}{2}$, for that is consistent with subsequent terms of $f(a)$ converging to 0 and consistent with subsequent terms of $f(a)$ converging to 1. However, as we cannot make it the case that $a > 0$ or $a \leq 0$ with a finite calculation with a pre-known upper bound on time consumption, there is no way to choose a third term of $f(a)$, for any possible choice would either be too far away from 0 or too far away from 1 to make it possible to have the sequence converge to that value if a subsequently attains a specific value (because a natural number is determined to have the property P or it is determined that it is impossible that any natural number does). Thus f is not a *total* function on the real numbers, but only a partial function defined for those real numbers that are either positive or non-positive.[42]

It would thus seem that the foundation of mathematics on mental constructions comes at a heavy price compared to Platonism. We are no longer allowed to reason using *tertium non datur* when we are not in a position to assert either the proposition in question or its negation. We are not permitted to use non-constructive proofs. Working with discontinuous real functions is also banned. And a further limitation of intuitionism, which has not yet been mentioned, is that many of the impredicative definitions that abound in classical mathematics are seen as illegitimate. An impredicative definition defines an object in terms of a totality of which it is an element. The classical mathematician is justified in using them, because the definition is merely a way to linguistically picking out an object which exists independently of the definition. However, for the intuitionist a definition provides a method of constructing the object which is constitutive of that object, and therefore the object is not guarantied to have existed as part of the totality in advance (in some cases they do: an impredicative definition of the form "the smallest natural number such that..." is in order, because any natural number can be constructed independently of the definition).

One important point that this dissertation makes is that the price of mentalism is much more modest than the intuitionist would have it seem. The reason is that the cost is to a large extend incurred because of the auxiliary doctrine of

[42]This was merely an illustration, meant to convey an intuitive understanding. It does not qualify as an outline of a proof, for then it would have to conclude by an application of double negation elimination, which is also intuitionistically invalid. A proper proof of the theorem proceeds from the Fan Theorem which is a corollary of the Bar Theorem to be discussed in Section 3.2. The simplest self-contained proof of the continuity theorem in the literature is, as far as I know, to be found in (Heyting 1956).

verificationism, not because of mentalism itself. The first part of this positive conclusion will be reached in Chapter 4. First, however, we must engage critically with Brouwer.

Chapter 2

Free choice sequences

A widely held view among the medieval logician-cum-theologians was that the omniscience of God implied bivalence for sentences about the future, even though this created problems for the doctrine of man's free will (Øhrstrøm 1984). So, prima facie it seems that Brouwer's denial of bivalence for a number of sentences about future constructions squares perfectly with his attempt to avoid making dubious metaphysical assumptions, such as the existence of a Platonic realm or, indeed, of God, to secure an ontological basis for mathematics. The aim of this chapter is to argue that this link is not as strong as generally believed, and that Brouwer's ontology can be combined with bivalence, or (see Chapters 5 to 7) at least with a logic where the counterexamples to bivalence are much more scarce.

2.1 Free choice sequences according to Brouwer

The one aspect of Brouwer's intuitionism that distinguishes it most from other types of constructivism is his use of choice sequences. A choice sequence is a sequence that is created in time by successive choices of new terms by a creating subject (Brouwer 1952, 142). Only a finite initial segment has been constructed, at any point in time. The sequence, therefore, is never finished, but always in a state of expansion. Thus, according to Brouwer, by basing mathematics on such objects, the need to assume that something actually infinite exists is avoided.

The subject can choose to pick the terms according to an algorithm; for example an algorithm that selects rational numbers which are increasingly better approximations to π. That brand of choice sequences are called *lawlike* sequences and are discussed in the next chapter. The opposite extreme are *lawless* sequences where each choice of term is made at random. The individual may grant himself the freedom of allowing each term to be *any* element of some species[1], e.g. the natural numbers, or he may elect, from the beginning of the construction or at any point during it, to impose restrictions on his own future choices. As long as these restrictions allow for more than one possible choice for each future term, the sequence is of a kind between the lawlike and the lawless. An important example is the decision to create a real number. This amounts to the subject imposing on himself the restriction that each term shall be a rational number q_n satisfying $|q_m - q_n| \leq m^{-1} + n^{-1}$ for all $m < n$.

Brouwer (1954) and Troelstra (1998, 199) claim that choice sequences *enforce* the use of intuitionistic logic. A simple example of where bivalence purportedly fails is given with a sentence that states that the 17th term of some lawless sequence is 99, when the sequence has only been developed to the 5th term. For a more interesting example, consider a sequence restricted so as to be a real number, but with no other restrictions, which so far only contains the three terms $-1/2$, $1/4$ and $-1/8$ (in that order). Here the sentence stating that the sequence is positive is neither true nor false according to Brouwer. For a real is, by definition, positive if one of its terms satisfies $q_n > n^{-1}$. The three terms constructed in the sequence so far do not secure that the sequence has this property and they do not secure that it does not. Likewise, it is not negative, defined as having a term such that $q_n < -n^{-1}$, nor equal to 0, which means that for all natural numbers n, $|q_n| \leq n^{-1}$.

2.2 Constitution of free choice sequences

So much for introductory explanations. We shall now turn on the critical sense and try to get a more precise answer to the question of what a lawless choice sequence is. What exactly constitutes it?

As is witnessed by the debate on personal identity, questions of constitution can often be elucidated by beginning with asking about the related questions of individuation and self-identity over time. So, if I begin a lawless sequence now at t_1 by making the first term 4, and then *now* at t_2 add 9 to it as its second term, what is it that makes the sequence at t_1 identical to the sequence at t_2?

The strongest possible answer, that they are qualitatively identical, can quickly be ruled out. If they were qualitatively identical they would have exactly the same properties, so if they were qualitatively identical, then the sequence should have the property at t_1 that it has 9 for its second term, as it has that

[1]A species is the intuitionistic counterpart of a set: an intensional collection of, possibly, potentially infinitely many objects. See Section 5.1 for a more detailed explanation.

property at t_2. So, by the same token, it would be the case for each n that at t_1 it would be a property of the sequence that there was some specific number that was its nth term. Then the sequence would be actually infinite.

Instead of the property being *has 9 for its second term*, it could be *has, at t_2 and later, 9 for its second term*. But this makes little difference because the problem still arises, *mutatis mutandis*, in that there are still an actual infinity of properties. The fact that some of them are *about* the future does not make for a relevant difference.

The failure of this attempt to reach a satisfactory answer teaches us two things: that we must look for some criterion of numerical identity instead, and that this criterion must allow for the sequence to be genuinely dynamic in nature. This is acknowledged by Brouwer (1955, 114) who wrote that:

> In intuitionist mathematics a mathematical entity is not necessarily predeterminate, and may, in its state of free growth, at some time acquire a property which it did not posses before.

However, commenting on this quote van Atten (2007, 14) states that:

> Observe that a property such as 'The number n occurs in the choice sequence x' is constitutive of the identity of x, but is generally undecidable and does not satisfy PEM [principle of the excluded middle].

If this were true, the property *the number 9 occurs in the choice sequence* α would be constitutive of α, which implies that the t_1-incarnation of α is not α. Consequentially, diachronic self-identity of a choice sequence would be impossible. At most, it can be the case that the property *the number n occurs in the choice sequence x* is constitutive of the identity of x from the point of time where n is added to the sequence. On pain of commitment to actual infinity, it cannot be before. And from that time onwards, it is decided.

In fairness to van Atten, it should be noted that, at this point, he might be thinking of the relation of equality instead of the relation of identity. And that is also an important distinction to make in order to make it clear that I am concerned with the latter and not the former. If I produce a choice sequence and you do so as well, following the law that for each term you pick the same number as I have chosen, our two sequences will be equal, but they will not be identical (at least, they won't be according to the conclusion reached by the end of this section).[2] The distinction between identity and equality is similar to the distinction between types and tokens. However, if equality is the relation van Atten has in mind, then I would argue that he is, within the quote above, guilty of smuggling an element of Platonism in via the back door, just as Brouwer does: van Atten makes a claim about what constitutes the identity of actually infinite routes, as will be discussed in Section 2.7.[3]

A second possible answer to the question of identity is that the identity of the sequence is grounded in a description such as "the sequence $\{a_n\}$ where a_n is the number thought of by person P at time n". The description is, in itself, a finite entity and can exist prior to the selection of terms by P, so there is no actual infinity in play in that respect. Of course, there are some rather banal problems with the exact formulation of the description, and therefore we have to do a little tweaking. It may not be determined in advance at what time the terms are chosen, so a description like "the sequence $\{a_n\}$ where a_n is the nth number thought of by person P after time t_0" is better. Then again, the creating subject may not have dedicated his life completely to creating the choice sequence in question and sometimes thinks of numbers which are not meant to be part of it. This problem can be solved by changing the description to "the sequence $\{a_n\}$ where a_n is the nth number which person P thinks of after time t_0 while having the intention that the number be part of the sequence". A final modification is needed; Brouwer claimed language to be external to mathematics. Mathematical objects are pure mental constructions of a single subject, so we cannot fix the identity of a choice sequence to a linguistic object like a description. But this is easily corrected. Instead we simply say that it is an idea of the subject corresponding to the mentioned description that serves to secure the identity of the sequence. This solves the more trivial problems connected with this proposal and leaves us free to turn to the more substantial ones instead.

The most substantial problem is that on any version of the description, each term is identified with what satisfies some definite description. This means that for the description of the sequence to be successful, it would seem that there must be a denoted object for each of the infinitely many instances of the definite description schema contained in the description of the sequence, in order that this object would have to exist at the time where the description

[2]Brouwer makes this distinction in his definition of "species" for example: "properties supposable for mathematical entities previously acquired, and satisfying the condition that, if they hold for a certain mathematical entity, they also hold for all mathematical entities which have been defined to be equal to it, relations of equality having to be symmetric, reflexive and transitive" (1952, 142).

[3]Reacting to a draft of this chapter, van Atten has informed me that he only intended to say that if the third term of α has been chosen to be 1, then it is known that a choice sequence β, for which something different from 1 has been chosen as its third term, is not identical to α.

is made. This does not in itself presuppose actual infinity, for terms can be repeated in a sequence and it may be the case that only a finite number of different terms appear within a sequence. However, Brouwer certainly does not want to restrict himself to such sequences. And whether or not there are infinitely many different terms, the actual infinite also threatens in another way. For even if 42 is both the 3th and the 5th term, it is two separate facts that *42 is the 3th number which person P thinks of after time t_0 while having the intention that the number be part of the sequence* and that *42 is the 5th number etc.*. So, it also seems that there has to be an infinity of facts about which terms satisfy the infinitely many definite descriptions.

Brouwer can of course not accept that these should be thought of as an actual infinity of different facts, for then he has merely reduced one kind of actual infinity to another. What will happen in the future cannot, in general, be facts in the present. That is, not when the assumption of the possibility of an infinite future is made, and Brouwer needs that premise. Hence, he is committed to anti-realism with respect to the future.

In general, the infinity of facts can also not be subsumed under a single, finite fact (nor any finite number of finite facts, for that is essentially the same thing, if we allow for conjunctive facts). If they can be described (or more Brouwerian: thought of) in a finite way, that would amount to a rule, and the very point of lawless sequences is that they are supposedly not all extensionally equal to lawlike sequences.

Apparently, the conclusion is that the definite descriptions of the future terms are not satisfied by anything. Therefore, the description of the sequence does not refer to anything. That means, there *is* no sequence. But now let us see what can be done to resist this conclusion. Consider the option that the description at any given moment, when n terms have been chosen, refers to an object of this form:

$$\langle a_1, \ldots, a_n, \_, \_, \_, \ldots \rangle$$

That is, the sequence's momentary incarnation is as an infinite sequence with n specific terms followed by an infinity of blank slots. The idea behind this rescue attempt for an otherwise doomed proposal is that actual infinity is avoided because just $n+1$ facts are needed; one for each of the fixed terms and one to the effect that all the remaining terms are "blank".

The notation "_" is suggestive but has to be substantiated. What we have been driven to is that it must signify some sort of middle ground, if that is possible, between the non-existence of the term and "full" existence of it. On the one hand, the term must exist enough to be able to fill the role of referent of the definite description, i.e. it must be the unique object satisfying the predicate "mth number thought of by person P after time t_0". On the other hand, the term can not exist in as strong a sense that it has the individual properties (like *being equal to 99*) which sets it apart from the other terms in the way that results in an actual infinity of distinct facts.

If such an intermediate form of existence can be allowed, this view of the constitution of the sequence seems compatible with Brouwer's claims about the failure of bivalence. The story would be that the properties which the "blank" terms may get in the future are not such that the terms do *not* have them now, in a strong sense of "not". Rather, it is undetermined whether they have these properties. As the sequence expands, there are no negative facts which change to positive in the way that more mundane change happens, e.g. when a tomato changes from being not red to being red. Instead, what is neither-nor becomes either true or false. The world becomes sharper, so to speak.

However, the price that a follower of Brouwer has to pay to defend his position in this way is obviously a heavy one. He is now committed to what could be called "Meinongism for parts",[4] i.e. that there are objects constituted by parts, some of which do not exist, but just subsist. Paying that price would defy the purpose of a philosophy of mathematics that is principally about avoiding dubious metaphysical assumptions. Subscribing to subsisting entities can hardly be categorized as metaphysically austere.

While the consequences for logic of this view agree with Brouwer's position, the view itself would be unlikely to gain his acceptance. He doesn't make positive ontological claims about parts of reality being undetermined, i.e. that there is a third ontological category between the existing and the non-existing. His rejection of *tertium non datur* is not backed up by a positive ontological story about objects with indeterminate properties; he makes claims of the form $\neg\forall x(P(x) \vee \neg P(x))$ but they are not backed up with claims of the form $\exists x \neg(P(x) \vee \neg P(x))$.[5]

Could a Brouwerian claim instead that even though the identity of the sequence is grounded in the definite description, some of the "contained" definite descriptions simply do not refer? No, for that answer makes it impossible to make certain semantic distinctions which Brouwer wants to make. On the one hand, he would claim that the sentence "all terms of α are natural numbers" is true, if the creating subject has imposed on himself the restriction only to

[4]Meinong famously suggested that there are non-existent objects in order to account for the meaningfullness of negative existential sentences and to secure a second relata for intentional directedness from a subject (Meinong 1904).

[5]See (Brouwer 1928), where x ranges over $[0;1]$ and P stands for the property of being rational.

select natural numbers as terms of the sequence α. On the other hand, the sentence "all terms of α are different from 99" is not true according to Brouwer, if 99 has not yet been chosen as a term. But, both the predicate "is a natural number" and the predicate "is different from 99" are satisfied by all existing terms of α, so they are not the cause of the difference, and terms that do simply not exist cannot do that work either. So the search for an extensional way to ground the identity of the sequence is misguided.

Let us therefore now move on to a third possible answer to the question of the constitution of a lawless sequence. The combined wisdom derived from the two first attempts is that we should be satisfied with numerical identity of the sequence over time, that the sequence should be understood as genuinely changing, and that not-yet-chosen terms should not be considered parts of the sequence. Does this then mean that an object of the form

$$\langle a_1, \ldots, a_n \rangle$$

is what we must conclude that a temporal instantiation of a lawless sequence is? No, for that is just an ordered n-tuple, and a choice sequence is obviously not just *that*.

There is a dynamical aspect to a sequence which is lacking from the n-tuple. This difference is, however, not in the past; also the n-tuple has been created, one term added at a time, in a temporal process. In Brouwer's universe there are no atemporal mathematical objects,[6] it is just that some of the temporal objects have been completed. That is the difference between the tuple and the sequence: the former has found its final form while the latter will continue to undergo changes.

This is, however, exactly the kind of claim that we have to be cautious about interpreting. The fact that it "will continue to undergo changes" must not be understood as an assertion about the actual future of the sequence, for the actual future does not exist. Given the commitment to anti-realism with regard to the future, the only content this claim can have is that the creating subject has an *intention* to amend the sequence. So, allowing "intention to expand" to be short for "intention to expand according to restriction R" if there is a such, this is our third proposal as to a constitution of a lawless sequence at a given instance of time:

$$\langle a_1, \ldots, a_n, \text{intention to expand} \rangle^7$$

[6]"[M]athematics [has] its origin in the basic phenomenon of the perception of a *move of time*, which is the falling apart of a life moment into two distinct things, one of which gives way to the other, but is retained by memory." (Brouwer 1954, 2, original emphasis)

[7]I may be slightly better to think of it as an ordered pair consisting of an n-tuble and the intention to expand: $\langle\langle a_1, \ldots, a_n \rangle, \text{intention to expand} \rangle$

This may be a simplification. Brouwer has been interpreted in phenomenological terms, according to which we do not experience extensionless points of time.[8] Terms are thus chosen during intervals of time rather than at instants of time. In addition to earlier terms being kept in retention and recollection, to use Husserl's (1964) terminology, the next term or the next few terms may be anticipated in protention. But the slight vagueness that this may introduce does not substantially influence the points that follow, precisely because there can only be a limited number of specific, individually chosen future terms which can be within the scope of protention. If an infinite number of terms is anticipated, it can only be in the form a rule, or simply as the anticipation of continuing to make choices (i.e. without the specific choices being part of the anticipation). So, simplified as it may be, this is our "official" third answer.[9]

And it is a very simple answer; the present product of an ongoing construction is merely what has actually been constructed plus the psychological fact that its creator does not consider it finished. The self-identity of the sequence over time does not rely on any objects in the future, but simply on the subject choosing, when he adds a new term, to identify the extended sequence with the old one.[10]

Brouwer seems to be assuming that the future terms are there in some sense, but are indeterminate in some respects. According to this answer, they are simply not there. That seems to be the answer which gets the stamp of approval of Ocham's Razor. Brouwer adds exactly the kind of mysterious, extra-mental entities to his ontology, which his project is all about eliminating.

Furthermore, there is no support for the assumption of there being indeterminate objects in Brouwer's official account of the mathematical ontology. And he is (what would otherwise count as) quite explicit in his delimitation of the mathematical realm: only mental constructs are admitted and only those that can be introduced in accordance with one of the two "acts of intuitionism" (Brouwer 1952).

The first act of intuitionism is the purification of mathematics, where everything that cannot be grounded in the intuition of time is exorcised. The intuition of time gives the subject the awareness of a difference in the form of the before-after relation, or in Brouwer's own words, the so-called Primordial

[8]See (van Atten 2007, 33–34) and (Becker 1923).

[9]On this proposal there is still, as in the second, intertemporal means to refer to the sequence, e.g. "the sequence started by person P at time such-and-such". The meaning of this description is constant even though its referent is changing.

[10]One might question whether such a decision to identify really has the force to secure actual identity. It does seem reasonable to me, but if I am wrong, the conclusion would instead be that there is no coherent notion of lawless choice sequences (given rejection of actual infinity).

Intuition of the empty two-ity. As explained in section 1.5, this can be translated into the numbers 1 and 2, and the number 3 can be created by holding one before-after relation in retention while distinguishing it collectively from a new "after". By repetition, the natural numbers can be constructed and so can any finite object or set of finite objects equipped with relations and operations in a way that is not much different from how it is done classically.

However, the second act of intuitionism is the realization by the creating subject that he is not limited to already created mathematical objects. Rather, he is free to employ the Primordial Intuition in any way he likes in a temporarily unbounded "free unfolding of the empty two-ity". This is what opens up for free choice sequences: the subject can set out to make a potentially infinite sequence consisting of "mathematical entities previously acquired".

The second act is what makes Brouwer's universe potentially infinite instead of finite. But, it is a potential infinity of Primordial Intuition-created entities. The second act does nothing to sanction new kinds of basic objects. It just allows for the open-ended addition and combining of more and more mental constructs.

2.3 Main argument

I will proceed on the assumption that the third answer is correct and argue that when there are no "shadowy" future terms, there are also no fuzzy properties to threaten bivalence, only what could suitably be named "unstable" properties.

Take as example the property of natural numbers that *the nth term of α equals 99*, when α is a lawless sequence that does not yet have 99 for any of its terms nor is under a restriction that prevents 99 from becoming a term in the future or guaranties the same. Calling this property P, the sentence $\exists n P(n)$ is not true according to Brouwer, and neither is its negation, that it is contradictory that there is an n for which P holds.

This is precisely what I want to dispute. To do this, it must first be noted that the notion of possibility which is used with the word "contradictory", is not the metaphysical one. If, for example, the two first terms of α have been chosen to be 6 and 7, he *would* consider the claim that those terms are equal to be "(known to be) contradictory", even though it is metaphysically possible that different terms could have been chosen, specifically chosen to be both, say, 6. Instead, it is epistemic possibility; "contradictory" means "not possible relative to what is known". Or to be closer to Brouwer's own formulations: that a sentence is contradictory means that the supposition that that sentence is true can be reduced to absurdity in a finite number of steps (Brouwer 1954).

But it is *not possible relative to what is known* that "at least one natural number n possesses the property that the nth term of α equals 99". For it is known that there are only a certain finite number of terms and that none of them equals 99. Say, for definiteness, that α at a given point in time is $\langle 6, 7, 13, \text{intention to expand}\rangle$. Then the supposition that *at least one natural number n possesses the property that the nth term of α equals 99* can be reduced to absurdity in the most straightforward way: The number 4 does not posses the property that the 4th term of α equals 99, as there is no 4th term. The same holds for all n larger than 4. So, knowing that there are only three terms of α, the supposition reduces to the proposition that either 6 equals 99 or 7 equals 99 or 13 equals 99, which is a disjunction of three absurdities and is therefore an absurdity itself.

And if a real number (i.e. a sequence in the process of being created according to the restrictions mentioned above) is defined to be positive if it has a term for which $q_n > n^{-1}$, then it is false that

$$\langle -1/2, 1/4, -1/8, \text{intention to expand}\rangle$$

is positive. For it is known that the sequence only has these three terms and that neither $-1/2 > 1$, $1/4 > 1/2$ nor $-1/8 > 1/3$ holds. When Brouwer claims that this sequence does not satisfy the criterion for being positive, nor the criterion for not being positive, this is not a genuine violation of bivalence. It is just a weird use of negation that is not forced upon us by the ontology of the subject matter.

What Brouwer claims are properties that violate bivalence, are just unstable properties. They are properties that a sequence may go from having to not having or *vice versa*. The mentioned sequence is not positive but will be if, say, 1/9 is added as the 4th term, 1/8 as the 5th, 1/7 as the 6th and 1/6 as the 7th. And in that case it would also change from having the property of being equal to 0[11] to not having that property. (Psychologizing a bit, it would seem that Brouwer's adversity to affirming of some object that it has a given property, if the object may lose that property later, is a residue of the belief, which he rejects, that all mathematical truths are timeless. He will allow that an object gains a property, but once it has gained it, it must be impossible for it to lose it again.)

This is the argument in a nutshell. I will attempt to support it by drawing on some insights from other areas of philosophy, namely the debates about definite descriptions, future contingents and fictional objects.

[11] This suggests that the definition of being equal to 0 should be changed, but that is beside the point.

2.4 Definite descriptions

If we look at the sentence "the 17th term of α equals 99" (assuming that less than 17 terms of α have been chosen) through Russell's eyes, the conclusion is easily reached: the sentence is equivalent to "there exists a unique 17th term of α and anything that is a 17th term of α equals 99", so due to the existence part, the sentence is false (Russell 1905).

The first reply one would think of would be that Brouwer could just side with Frege instead in the famous "the present king of France is bald" controversy. According to Frege's analysis, that sentence is neither true nor false (Frege 1892), and the same holds for "the 17th term of α equals 99". But that would not work for Brouwer; Frege's theory does not in general support Brouwer's claims about the truth values and lack thereof of mathematical statements. It is a consequence of Frege's theory that any sentence consisting of a predicate being applied to "the 17th term of α" is without truth value, while Brouwer would have it that a sentence such as "the 17th term of α is a natural number" could be true if the proper restriction had been decided on. It is a common problem of the theories of Russell and Frege, seen from the Brouwerian point of view, that they result in over generation of non-true sentences.

This again leaves the Brouwerian with just Meinongism as a general theory of language under which Brouwer's view of the semantics of mathematical propositions could be subsumed: the future terms exist partially so as to have some properties "decided", i.e. either true or false of the term, and others undecided. But of the three theories of definite descriptions, it is only Russell's that doesn't rely on metaphysical assumptions which are unacceptable to Brouwer. Meinongism is an incarnation of Platonism, and Frege relies on *ein drittes Reich*, in addition to the worlds of physical and mental objects, inhabited with timeless *Gedanken* and *Sinnen*.

So, it would seem that Brouwer would have to surrender to Russell's analysis of sentences like "the 17th term of α equals 99" and admit that they are true or false. An objection that has to be considered is this one: Brouwer's notions of truth and falsity are epistemic not ontological. To be "true" means to be known or proved (more on this in the next chapter). Ergo, the objection goes, the discussion in the last few paragraphs is misguided as it focuses on the ontology of the choice sequences.

That objection is not difficult to shoot down. The analysis of the constitution of a lawless sequence reveals that it is an epistemically transparent object. As it only consists of the already chosen terms plus the intention to carry on, the creating subject knows everything there is to know about it.[12] There are no unknown facts about future terms. They are not there, and the subject knows just that. Therefore, in this case, ontology and epistemology comes to one and the same thing.

A more interesting objection is perhaps that the problem with the discussion is not its focus on ontology, but its focus on language. According to Brouwer, language is merely an imperfect means for communication of mathematics which in itself consist in languageless mental constructions (Brouwer 1907, 73, 79). So, it seems rather odd to mount a critique of his views based on the semantics of definite descriptions. My reply is as follows: what we should presumably analyze instead of the *sentence* "the 17th term of α is 99" is the creating subjects mental state of *intentional directedness* towards the future 17th term of α and his *belief* that it equals 99, and I do not see how that could make any difference; the belief is simply false. Of course, my lack of imagination does not constitute a very strong argument, so let us just phrase it as a challenge to the Brouwerian: what are the relevant differences between language and thought which implies that Russell's analysis cannot be carried over?

2.5 Future contingents

However, there is at least one genuine problem present within this discussion of definite descriptions, namely that it is not sufficiently general. A sentence such as "99 will be added to α" does not have a definite article, but the problem is the same. The core of the problem is sentences about the future. In the debate about future contingents, theories have been suggested that imply that certain sentences about the future are neither true nor false (Barnes and Cameron 2011). So we should examine these theories to find out if they may support Brouwer.

Prominent in the discussion of time is the idea of *branching time*: from the present, there is "access" to several different possible versions of how the world is in, say, one second from now, and the path to each of those split up in different possibilities for the world as of two seconds from now. If in each of the paths that can be taken through such nodes from the present into the indefinite future, 99 is at some point added to α, then the mentioned sentence is true; if 99 is not added to α in any of the paths, then the sentence is false;

[12] At least he knows the "simple" facts such as what the terms are. A proposition such as "the first term of α equals the first digit to follow the first occurrence of one hundred consecutive 0's in the decimal expansion of π", which is really about more than just the lawless sequence in question, is another matter.

and if 99 is added to α in some but not all the paths, then the sentence is neither true nor false.

There are two ways to interpret this idea of the future as a branching tree. One is realistically, i.e. to claim that the nodes represent existing possible worlds. The passage of time is then likened to the movement of a light beam, which traces a path through the tree. With that interpretation, it is easy to account for the semantics of sentences about the future, for truth, falsity and neither-nor can be explained as correspondence or lack thereof to these worlds. However, as already argued, realism about the future is not a viable way out for Brouwer, so realism about a plurality of futures certainly is not either.

The second interpretation is that the idea of branching trees is just a tool for making our discourse and reasoning about the future more precise. That is how talk of possible worlds in general is understood by most philosophers. But, in that case, the idea is irrelevant for our present purpose. We are looking for a metaphysical story to back up the claim of failure of bivalence, and this second interpretation amounts to taking the metaphysical commitments out of the theory. So, we must seek answers elsewhere.

An alternative theory about the future which also results in the rejection of bivalence is that there is just one future, but that it is indeterminate in some respects. So for example, if I am now standing under a cloudless sky in Aberdeen, the one-second-from-now future is determined with respect to the sentence "the sky over Aberdeen is not completely cloud covered" being true, while the sentence "Queen Elizabeth is moving her right arm" is neither true nor false, due to Queen Elizabeth having a free will, so the future is indeterminate in that respect.

The question that arises from this is how to interpret this indeterminacy. There seems to be three possible ways to understand the concept of indeterminacy: either semantically, epistemically or metaphysically (Torre 2011). The truth value of a sentence is semantically indeterminate if the indeterminacy is due to a word in the sentence not being defined sufficiently precise – if it is vague. Brouwer was definitely skeptical about our concepts being as sharp as we generally think they are (Brouwer 1905). Nevertheless, it is certainly not his contention that "99 will be added to α" lacks truth value for *that* reason. Because then a shift from focusing on language to focusing on thoughts of an ideal mathematician would remove the indeterminacy.

Epistemic indeterminacy would in this case mean that the absence of a truth value is because of the lack of knowledge about α. We have been at this point before and rejected it. This interpretation would imply that there is something to be ignorant about. But, being a creation of the free mind of the subject, there *is* only the part of α that he knows about. A lawless sequence does at t not transcend the creating subject at t. So, the option of epistemic indeterminacy reduces to metaphysical indeterminacy. That option is the position that the future exists and some parts of it are sharp while other parts are blurry. And the Brouwerian judgment about this must be the same as about the realism concerning a branching future, namely that it is far too metaphysically "heavy".

We must stick to anti-realism about the future and investigate what follows from that concerning the semantics of future-talk. Dummett (2006, 19), using the term "presentism" for anti-realism about the future (and the past), claims that "presentism would necessarily require a semantics that repudiated the principle of bivalence". I will contest the claim to necessity, but grant that presentism *can*, if one is willing to accept an unpleasant consequence, motivate the rejection of bivalence. If we take correspondence with the future to be the truth criterion of sentences about the future, then it might seem reasonable to stipulate that they are neither true nor false, because the future's status as something that does not yet exist but will, warrants a semantic treatment of it different from that which does simply not exist. But then consistency requires that we evaluate *all* sentences about the future in that way. So going back to the example of the sentence "the 17th term of α is a natural number", we would have to declare it to be without truth value because that part of the future which contains the 17th term does, as the rest of the future, not exist. So, in this case we also get a result which the Brouwerian considers over generation of non-true sentences.

Instead, we should follow Peirce (1935, §368) in locating the truth makers of sentences about the future in the present. So the sentence "the sky over Aberdeen is not completely cloud covered in one second from now" is true due to the present Aberdeen being in such a state that huge amount of clouds cannot form over it in just one second.

This has consequences for the issue of which modalities are available for future-talk. Prior (1967, 128–136) attributes to Ockham (1969) the answer that there are three: necessity, factuality and possibility (or rather just two, for possibility can be defined in terms of necessity in the usual way). He claims that it might be the case that it is possible that it will rain tomorrow and possible that it will not, while it is neither necessary that it will rain nor that it will not, and that yet the sentence "it will (as matter of fact) rain tomorrow" is true. But, if we stick to sentences about the future actually being about the present, then we cannot uphold a distinction between something actually happening tomorrow and it necessarily happening.[13] Therefore, we must instead follow Prior's "Peircean" doctrine that all factuality claims about the future are necessity claims in disguise.

So "99 will be added to α" is to be read as "necessarily, 99 will be added to α", which is false. And its apparent negation, "99 will not be added to α" means "necessarily, 99 will not be added to α" which is also false, and is so without violating any principle of classical logic.[14] This interpretation gives Brouwer the desired truth value of "the 17th term of α is a natural number"; read with an implicit necessity operator, it is true by virtue of the presently existing restriction on future terms of α. That is, "the 17th term of α is a natural number" means "the restriction prescribes that when a 17th term is chosen, it must be a natural number", and it is the restriction that supplies the truth maker and not a 17th term.

Of course, nothing that has been said here precludes a Brouwerian from acknowledging both that there is no future and that sentences about the future must be understood as being about present necessities, and yet choose to stipulate that such a sentence should count as neither true nor false when the content of the sentence is neither necessary nor impossible. However, we are now very far from this choice of semantics being *forced* upon us as claimed by Brouwer, Troelstra and Dummett. To the contrary, it seems like a rather unreasonable introduction of indeterminacy into the language about something (the present) which is itself determinate.

Such a stipulation is exactly what the intuitionist makes.[15] *Qua* stipulation I can see no objection to it, but the intuitionist couples it with a claim that there is no stronger, legitimate notion of truth and falsity, and that claim is objectionable. The claim is that I cannot truthfully state the negation of "99 is a term in α" because I cannot rule out that this proposition can be proved, and the reason that I cannot rule that out is that a proof of the proposition is a construction which in finitely many steps provides an n such that the n'th term of α is 99, and, furthermore, such a construction may include *making more terms of* α.[16] But, that compares to the situation where I am asked

[13] That is unless we take the actuality-statements to be true when a certain threshold of probability is exceeded or if there is necessity just with the exception of *ceteris paribus* possibilities. That may be a reasonable way out in the case of the weather examples, but it is not relevant for choice sequences.

[14] A sentence such as "it is not the case that 99 will be added to α" is ambiguous between a wide-scope and a narrow-scope reading, i.e. it can either be taken to mean "it is not the case that necessarily, 99 will be added to α" or likewise mean "necessarily, 99 will not be added to α". But such ambiguity is simply to be resolved by stipulation.

[15] In Section 3.5 it will be formulated with a little more precision exactly what the stipulation is.

[16] Thanks to van Atten for pointing this out.

what the likelihood is that it will rain tomorrow and I then initially refuse to answer the question, wait a day, get wet and answer "100%", thereby answering a different question from the one I was asked. There is an answer I can give to the question of rain on the day that it is asked, based on existing meterological facts, and likewise there is an answer to the question whether 99 is a term of α that I can give *now*, namely "no". Given the conclusion in Section 2.2, a refusal to give that answer constitutes a vow of silence, not a violation of classical logic.

To sum up, in combination, the points about temporality and about ontology versus epistemology, when van Atten et al. (2002, 214) claim that

> This unfinished character of choice sequences has repercussions for logic. It means that a sequence can not, at any stage, have (or lack) a certain property if that could not be demonstrated from the information available at that stage. It follows that bivalence, and hence the Principle of the Excluded Middle, does not hold generally for statements about choice sequences. For example, consider a lawless sequence α of which so far the initial segment (1,2,3) has been generated, and the statement P = 'The number 4 occurs in α'. Then we cannot say that $P \vee \neg P$ holds.

they are wrong. Taking the present tense of P literally, P is just false. And changing P to the future tense, "The number 4 will occur in α", we cannot interpret the sentence at face value, for

> there are things about the future that God doesn't yet know because they're not yet there to be known, and to talk about knowing them is like saying that we can know falsehoods. (Prior 1996, 48)

Instead, it must be understood as a claim about the present necessitating that 4 will be added to the sequence and, as such, be either true or false depending on which restrictions have been placed on α. (For obviously, necessity cannot be analyzed extensionally as being about possible worlds, but must be grounded in the intensional properties of the sequence.)

As an aside, before moving on to considerations about fictional objects, let us consider an observation of Posy's (1976): there are two different notions of negation in play in Brouwer's writings. Posy refers to the following example, which is about a lawlike sequence, from (Brouwer 1924b): For all natural numbers n, define c_n to be equal to $(-1/2)^{k_1}$ if k_1 is smaller than n and the sequence 0123456789 appears for the first time in the decimal expansion of π with the 0 at the k_1'th decimal position, and equal to $(-1/2)^n$ if there is no such k_1. Then define r to be $\langle c_1, c_2, c_3, \ldots \rangle$. Concerning this number Brouwer writes (p. 252) that “die Zahl r ist nicht rational, trotzdem ihre Irrationalität absurd ist”.[17]

Just a few lines earlier, he has defined the irrationality of a number to be the absurdity of its rationality. Hence, if we equate the negations expressed by “nicht” and “absurd” then the quoted sentence asserts the conjunction of the negation and the double negation of one and the same proposition. So we have to distinguish them. Posy uses the symbol $\neg$ for the intuitionistic absurdity and the symbol $\sim$ for the simple denial.

He then goes on to propose a reconstruction of Brouwer's logic in terms of a reduction of $\sim$ to $\neg$. He does so using a knowledge operator: $K_n^a P$ means that at time n agent a knows P. The claim is that $\sim P$ can be identified with $\neg K_{n_0}^a P$, where n_0 is the present. With this interpretation some central Brouwerian doctrines are salvaged. First, we do not have $P \vee \neg P$ in general, just $K_n^a P \vee \neg K_n^a P$. And second, no proposition ever changes between being true and false, for if a does not know P at some time n but learns it later at m then $\neg K_n^a P$ is true at n and stays true, $K_m^a P$ becomes true (i.e. changes from not having a truth value to being true) and P becomes true.

Considered as a purely exegetical thesis, I can only endorse it. My *systematic* point is that the reduction should go in the opposite direction: $\neg P$ should be defined as $\Box_n \sim P$ meaning that at time n it will necessarily be the case in the future (in the sense explained above) that $\sim P$. Then the principle that is refuted by choice sequences is just $\Box_n P \vee \Box_n \sim P$ which is not a law of classical logic but is, nonetheless, a principle that the Platonist believes holds for all mathematical propositions. Then, of course, the question is whether lawless choice sequences should be admitted as mathematical objects. That question we return to in Section 2.7.

2.6 Fictional objects

A likeness between fictional objects and lawless choice sequences is another thing that may induce one to claim that bivalence fails for the latter. For it is natural to think that it does for the former. The idea is that since Shakespeare does nowhere in *Hamlet* specify Hamlet's height, Hamlet is an incomplete object not having a specific height, so a sentence such as “Hamlet is 5 foot 7” is neither true nor false. And it seems that the crucial feature of fictional objects which is responsible for this incompleteness is that they are creations of our minds. Since choice sequences share this feature, the analogy argument goes, they too are incomplete and defy the principle of bivalence.

[17] “The number r is not rational, in spite of its irrationality being absurd”.

I think it is possible to develop a convincing argument against this conclusion which has the same structure as in the previous section, namely by going through the various different accounts that there are of fictional objects and showing that only those according to which bivalence does hold, are metaphysically acceptable to a Brouwerian. This would be somewhat repetitive, however. As an alternative argumentative strategy is available, I will only focus on that. I will grant, for the sake of argument, that fictional objects are incomplete, and instead attack the claim that there is a relevant analogy.

In (van Atten 2007, 95) where van Atten draws the analogy, he notes that there is one important difference. While "a choice sequence is something we literally bring about in the process of its construction", a fictional object is just a piece of make-believe. When Brouwer's creating subject thinks of a new choice sequence and imagines that it has 4 as its first element, then a choice sequence is actually created and it is literally true to say that it has 4 as a term. But, when Shakespeare writes about Hamlet and informs his readers that he is a prince of Denmark, Hamlet does not spring into existence (at least not as the actual person he is according to the story, but one might follow Thomasson (1999) in saying that Hamlet qua fictional object is created in that instant) and it does not become true that Hamlet is a prince. This difference really is crucial and is more so than van Atten appears to acknowledge. The only aspect of this difference that he takes note of, is that whenever a property is indeterminate of some choice sequence, that property may later be determined, while the indeterminacy of fictional objects is lasting.

The indeterminateness of fictional objects steams from a certain tension. We evaluate sentences about fiction pretending we are in a world different from our own. On the one hand we think of this world as a complete world, just as our own. *Hamlet* is not a weird science fiction novel about an incomplete world. It is about a world which is metaphysically like our own, with just a few facts about what goes on in Elsinore changed. On the other hand Shakespeare has not made it complete, i.e. he has not supplied us with a truth value for every possible proposition about it. So while the proposition that Hamlet has a height must be true according to the fiction, because Hamlet is portrayed to be a normal human being, there is no truth value for the proposition that Hamlet is 5 foot 7, nor any other proposition attributing a specific height to Hamlet.

The objects of fiction are incomplete in the sense that they do not have a "full set" of properties. However, this is essentially a comparative claim. What a full set is, depends on real-world objects. It is only because actual human beings have a height that there is something for Hamlet to be missing.

Now try to draw the analogy between fictional objects and choice sequences with this in mind. The result is that the incompleteness of a choice sequence not yet having a specific number as its 17th term owes the status of being incomplete to a comparison with some real sequences. But, it is of course not an acceptable consequence for a Brouwerian that we must postulate the existence of some sequences that are more real than those that are mental constructions. This is especially because they must then not only be characterized by all having a specific 17th term, but also by having a specific n'th term for every n. That is, they must be actually infinite.

Not having a specific 17th term is only an incompletness in the same sense as Hamlet's not having a specific height, if the proposition that the sequence has a 17th term follows from being a sequence as the proposition that Hamlet has a height follows from his being a human being. However, it does not according to the intuitionistic conception of sequence. It only follows that if a sequence does not yet have a 17th term then there is an intention to create one at some point in the future. The idea that all real sequences have an infinity of terms seems to be another residue of classical doctrine just as the unwillingness, noted above, to allow for the truth value of a sentence changing between true and false.

We must conclude that the analogy is not valid, at least not for the purpose of arguing for the failure of bivalence. When fictional objects and choice sequences are considered alike by being incomplete, it is a conflation of two very different senses of "incomplete": not having "as many" properties as real objects and not having stopped changing.

2.7 Consequences for the continuum

Let us investigate the consequences of our conclusion for the concept of real numbers and our understanding of the continuum.

In his early years, before he came up with the idea of free choice sequences, Brouwer was of the opinion that the continuum is a primitive notion. It cannot be constructed out of entities of another type. Specifically, it can not be identified with a set of points. The continuous and the discrete are described as complementary and equally basic aspects of the Primordial Intuition (Brouwer 1907). Points (and numbers) can only be used to analyze a pre-given continuum by being the endpoints of the subintervals into which it can be decomposed.[18] One reason that a continuum cannot be a set of points is that

[18] This is the view originating with Aristotle (1930).

the available points are only those that can be identified with rational numbers or definable real numbers, i.e. lawlike sequences, which implies that there is only a denumerable infinity of them and hence not enough to exhaust the continuum (Brouwer 1913).

The intention with lawless choice sequences is that they can improve on the situation:

> [Intuitionism] also allows infinite sequences of pre-constructed elements which proceed in total or partial freedom. After the abandonment of logic one needed this to create all the real numbers which make up the one-dimensional continuum. If only the predeterminate sequences of classical mathematics were available, one could by introspective construction only generate subspecies of an ever-unfinished countable species of real numbers which is doomed always to have the measure zero. To introduce a species of real numbers which can represent the continuum and therefore must have positive measure, classical mathematics had to resort to some logical process, starting from anything-but-evident axioms[...] Of course, this so-called complete system of real numbers has thereby not yet been created; in fact only a logical system was created, not a mathematical one. On these grounds we may say that classical analysis, however suitable for technology and science, has less mathematical reality than intuitionist analysis, which succeeds in structuring the positively-measured continuum from real numbers by admitting the species of freely-proceeding convergent infinite sequences of rational numbers and without the need to resort to language or logic. (Brouwer 1951b, 451–452)

There are two ways to interpret this. The stronger interpretation is that Brouwer now does, in one crucial respect, exactly the same as the classical mathematician, namely finding a non-denumerable totality of points with which to identify the continuum.[19] True, there are still major differences in what that totality of points looks like, but the extensional conception of the continuum as a collection of points is now adopted by Brouwer. Let me first explain why Brouwer is wrong, under the assumption that this is indeed his postulate.

In addition to the alleged indeterminate character of Brouwer's reals which results in the continuum not being completely ordered (the disjunction of $x < y$, $x = y$ and $x > y$ does not hold for every pair of real numbers x and y), one other difference should be noted. Where the classical reals (interpreted Platonistically as Brouwer does) by virtue of the hierarchical nature of the set

[19] Or rather, since Brouwer disputes that the classical mathematician succeeds and I dispute that Brouwer does, the formulation must be more careful: what Brouwer believes he does is the same as what the classical mathematician believes *she* does.

theoretical universe must all exist for the set of them to do the same, Brouwer only commits himself to the *possibility* of constructing each of his reals. They do not all have to exist prior to them being collected in the species of all reals. His continuum is the totality of all possible convergent sequences of intervals.

This difference notwithstanding, my claim is that Brouwer's approach fails. Here is why. Our analysis of the constitution of choice sequences shows that there are not uncountably many of them. For we can not individuate them by how the entire process of choices goes. They can only be individuated by who the creating subject is, starting time and, at any given time, the created initial segment and adopted restrictions. Hence there are always just finitely many of them, i.e. the class of choice sequences is potentially, countably infinite.

At first sight, this objection to Brouwer may seem to ignore the just mentioned difference between Brouwer's and the classical reals. So, the Brouwerian might answer: yes, we can only ever have initiated the construction of a finite number of lawless choice sequences, but there are non-denumerably many different ways they might go – non-denumerably many different possible routes, so to speak.

However, this is incorrect. The actual infinite, non-definable sequences (assuming for the moment that they exist) do not correspond to possible routes for potentially infinite choice sequences. For "possible" means "can be taken", and the *entire* route corresponding to a Platonistic non-definable sequences can never be taken. Only initial segments of those sequences can ever be taken.

This may get clearer if we make use of a nice metaphor of Posy's (1976, 98–99). He likens choice sequences to the route of a bus traveling on a forking highway. The journey of the bus can be seen from different perspectives. First, there is the perspective of a passenger in the bus seated with his back to the driver so that he can only see the route already traversed. Second, there is the perspective of the bus driver which in addition to the knowledge of his passenger has an intention of where to travel from his present position. Third, and last, there is the perspective of a pilot looking down on the bus and the road system from a helicopter hovering above, seeing both the traveled path and the roads ahead.

Given the rejection of actual infinity there is no helicopter perspective. For lawless choice sequences there are no roads before they have been traveled. Actually infinite roads are no less actually infinite than completed infinite travels. The only legitimate perspectives are the passengers and the drivers and in the former case that is a finitely extensional perspective and in the latter the perspective is finitely extensional and finitely intensional.

For the bus driver or the creating subject, there is an infinity of possibilities. But one must not conflate an infinity of possibilities with the possibility of infinity. The conclusion is that choice sequences cannot do the same work as the classical set of real numbers allegedly do.[20]

That brings me to the weaker interpretation of the quote. In the strong interpretation, I assumed that Brouwer's claim was that his reals make up the one-dimensional *intuitive* continuum. However, an alternative reading is that they just make up the *mathematical* continuum, i.e. that they make up the best possible model we can have of the intuitive continuum. That is consistent with this model falling short of being a perfect one-to-one model. With this interpretation, Brouwer makes a more modest claim, namely that the lawless sequences adds to the model something which the lawlike sequences cannot accomplish.

However, our analysis in the previous sections reveals that also this more modest thesis is mistaken. The addition of lawless sequences does not improve on the mathematical treatment of the continuum. For notice that for a choice sequence, which is a real number, when n terms have been chosen, all the information of those n terms can, without loss, be summed up as an interval to which the limit of the sequence is now restricted. Therefore, it makes no difference for the theory of real numbers if we identify the development

t_1: $\langle -1/2, \text{intention to expand}\rangle$
t_2: $\langle -1/2, 1/4, \text{intention to expand}\rangle$
t_3: $\langle -1/2, 1/4, -1/8, \text{intention to expand}\rangle$

with

t_1: $\langle [-3/2, 1/2], \text{intention to change}\rangle$
t_2: $\langle [-1/4, 1/2], \text{intention to change}\rangle$
t_3: $\langle [-1/4, 5/24], \text{intention to change}\rangle$[21]

At any given time, the mathematical content of a lawless sequence equals an interval with rational endpoints. The creating subject is just changing his mind about which interval to use, and, if it wasn't for the rather arbitrary details of the definition of "real number", each choice is one that could have been made initially. The implication is that free choice sequences do nothing that rational numbers cannot do.

[20] Also Brouwer's notion of "spreads" can only be understood in two ways: either Platonic, as a completed tree of possible routes, or Aristotelian, as a potential infinity of finite ways to subdivide an interval. There is no middle ground.

[21] The interval for t_3 is this intersection: $\left[-1/2 - 1^{-1}, -1/2 + 1^{-1}\right] \cap \left[1/4 - 2^{-1}, 1/4 + 2^{-1}\right] \cap \left[-1/8 - 3^{-1}, -1/8 + 3^{-1}\right]$; similarly for the interval for t_2.

Add to this that they have rather strange logical properties: *tertium non datur* doesn't fail, but "necessity-*tertium non datur*" ($\Box_n P \vee \Box_n {\sim} P$) does and some properties are unstable. Then it seems that it would be best to avoid free choice sequences in mathematics. One way to do this would be to deny that they are legitimate mathematical objects. However, I do not think that is a reasonable move. Assuming a commitment to mathematical objects being mental and temporal (which we will make in Chapter 4), it would be very difficult to draw the demarcation line between the kind of objects we would like to include and lawless choice sequences in a principled way.

Instead, I suggest that we just ignore the lawless sequences and only consider law-like sequences when we do analysis. And if the Brouwerian questions our right to do so, then *he* has a demarcation problem. For if free choice sequences have to be admitted into analysis, what is to rule out free choice-and-change-of-mind sequences? The creating subject can embark on creating a sequence, choose the first ten terms and then replace the first term with something else. And he can claim that the sequence is self-identical over time with the same right that he can claim that an ordinary free choice sequence is, for, recall, that identity is based on nothing but the subjects intention to consider them identical. Then hardly any property would be stable, which means that if we follow Brouwer in only allowing ourselves to affirm stably true sentences, then we cannot state that "6 is the 1st term" even if it is (at present) and we can never affirm of a free choice sequence that it is a positive number. The Brouwerian *decides* to not quantify over free choice-and-change-of-mind sequences when he says "all real numbers" (otherwise real numbers would not have the essential property of being arbitrarily approximatisable), so we can *decide* not to quantify over lawless choice sequences.

Chapter 3

Brouwer's conception of truth

Let us now turn to lawlike sequences. Brouwer claims that bivalence also fails for these sequences. A lawlike sequence is a potentially infinite sequence where the creating subject has imposed such strict restrictions on his own future "choices" of terms that there exists exactly one possible choice for each term. This existence claim has to be understood intuitionistically – so the restriction must be formulated in such a way that it is effectively possible for the subject to calculate the sole allowed term. That is, the restriction must take the form of an algorithm. However, let us follow Brouwer in calling it a "law".

In order to illustrate Brouwer's claim, we can use the example mentioned at the end of Section 2.5, i.e. the sequence $\gamma = \langle c_1, c_2, c_3, \ldots \rangle$ where the procedure for deciding c_n for a given n is as follows: Calculate the first $n + 9$ decimals of π. If the sequence 0123456789 does not appear anywhere, let c_n be equal to $(-1/2)^n$. If it does, let c_n be equal to $(-1/2)^{k_1}$, where k_1 is the position of the digit 0 in the first such appearance.

Brouwer's claim is that the proposition that the real number γ equals 0, is neither true nor false. It would only be true if we had a proof that there are no 0123456789-sequences in the decimal expansion of π, and it would only be false if we knew of such a sequence.[1] Assuming that it is true or false in the absence of such knowledge amounts to assuming that the decimal expansion of π has extra-mental existence.

The purpose of this chapter is to provide an interpretation of what exactly Brouwer means by "true" when he makes this claim. That is, this chapter

[1]We actually do today: there is a 0123456789-sequence beginning at decimal number 17,387,594,880 as well as at several other, later positions (Wells 1986). But as Brouwer (1951a) correctly points out, there will probably always be an ample supply of other examples that can be used instead. Hence, this is not really relevant, so we will just stick to Brouwer's example.

is purely exegetical (or perhaps more accurately: is an attempt at a rational reconstruction). In the next chapter, this interpretation will be used as a basis for criticizing Brouwer and for suggesting an alternative.

Before I present what I think is the correct interpretation of Brouwer, I will discuss and reject three other possible interpretations. I do this not just for the fun of shooting them down, but because my interpretation combines elements from those three and is, therefore, best understood and motivated against the backdrop of these erroneous alternatives. The first interpretation is that what is true is only that which has actually been constructed. The second is that not only actual but also all potential constructions can serve as truth makers. (These first two alternatives are quite naïve and are not serious contenders but they serve a purpose of stage setting.) And the third is that truth is equated with proof. This third alternative has now become so entrenched as an interpretation of Brouwer that a word of warning is in order so as to forestall misunderstandings where it is read into the first two: The notion of proof plays no role in the first two interpretations and "construction" is not to be read as "construction of a proof". Rather, "constructions" are of the mathematical objects and relations between them that mathematical theories are about, not of the proofs and theorems about them.

So much for introduction. Now we can get down to business and ask the question of what notion of truth Brouwer applies as his alternative to the rejected Platonic notion. In several places he gives rather explicit answers to this question – answers which are nonetheless puzzling. Here is one:

> *[T]ruth* is only in *reality*, i.e. in the present and past experiences of consciousness. Amongst these are things, qualities of things, emotions, rules (state rules, cooperation rules, game rules) and deeds (material deeds, deeds of thought, mathematical deeds). But expected experiences, and experiences attributed to others are true only as anticipations and hypothesis; in their contents there is no truth. (Brouwer 1948, 1243, original emphasis)

So Brouwer's official theory is that truth consists in correspondence with actual constructions. However, at first sight it does not seem that he adheres strictly to this credo. One thing that seems to be at odds with it, is his claim that as soon as the subject has a decision procedure for a given proposition then *tertium non datur* holds for it. Sticking to actual constructions as truth makers, one should presumably say that only when the decision procedure has been executed does the proposition gain a truth value. Another thing is that the theorems that he states are typically like classical theorems in that they cover an infinity of cases even though, obviously, not all these cases can have been realized as actual constructions. So, it seems that Brouwer often relies on potential rather than just actual constructions.

If that is the case then it appears easy to explain why Brouwer has felt the pressure to do so. The claim that the "actualist" position commits you to, that the proposition that, say, the one millionth decimal of π is 1, was not true at Brouwer's age but only has become so since, is very weird. It is weird because, by Brouwer's own admittance, it is determined in advance what the decimals of this lawlike sequence are.[2]

So is the theory of truth which Brouwer actually believes in, that whatever is determined in advance is true, i.e., that predetermined potential constructions are sufficient for truth? No, Brouwer does not go nearly far enough in this direction to warrant such an interpretation.[3] For it is also determined in advance whether there is a k_1, and if there is, what value it has. So by that standard it would be false, atemporally, that the limit of γ equals 0.

Framing the same point in terms of the popular example of Goldbach's Conjecture, it cannot be the case that it is fixed in advance for each n whether it would be a counterexample to the conjecture, but not fixed in advance whether such an n can be constructed. If we let P stand for the property of being a Goldbach number and n range over the even integers greater than 2, Brouwer claims that $\forall n(P(n) \vee \neg P(n))$ is true but that $\forall n P(n) \vee \neg \forall n P(n)$ is not. That difference can not be accounted for if truth is a matter of predetermined potential constructions alone, for the two propositions are about the same potential constructions.[4]

Consistent reliance on potential constructions would make all truths about lawlike sequences timeless and independent of the subject's knowledge and would therefore be in conflict with the temporality of Brouwerian truth and in particular with the role played by possession of algorithms; when the subject acquires a means to "judge" a proposition, i.e. comes up with a decision procedure for it, *tertium non datur* becomes valid for it (Brouwer 1952, 141).

[2]Brouwer also uses the word "predeterminate" for "lawlike" (Brouwer 1948, 1237). He also writes that the "freedom in the generation of [a free choice sequence] may at any stage be completely abolished [...] by means of a law fixing all future [terms] in advance" (Brouwer 1954, 7).

[3]Even though he also commits explicitly to this second interpretation in writing, like he did for the first: in his own copy of his dissertation he changed "*bestaan* in wiskunde betekent: *intuïtief zijn opbebouwd*" to "*bestaan* in wiskunde betekent: *intuïtief op te bouwen*", that is, "*existence* in mathematics means: *to have constructed intuitively*" to "*existence* in mathematics means: *to be constructible intuitively*" (van Dalen 2001, 134, footnote f).

[4]One way to analyze the concept of predetermination is with a counterfactual: if at time t_2 the nth decimal of π is found to be m, then for any $t_1 < t_2$ it would have been the case that if an agent had constructed the nth decimal at t_1, it would have been m. Does Brouwer deny this as the Kripkensteinian rule-following skeptic (see Section 4.3) does? I don't think he does. He just refuses to recognize such facts of predetermination as truths, presumably because he cannot locate a truth-maker for it within his anti-realist ontology.

Neither actual constructions nor potential constructions can be made out to be Brouwer's criterion for truth. Rather, he seems to be somewhere in between, relying on potential constructions when the subject has knowledge of its finitude in advance and otherwise insisting on actual constructions. That difference can not be explained with mentalism alone. So if this intermediate position is to be seen as more than an arbitrary compromise between conflicting sources of pressure, there must be some more fundamental truth criterion in play which can explain the unequal demands on what kind of existence of constructions is required.

Partly as an answer to this challenge, it has become common to interpret Brouwer as equating truth with existence of proof, or as it is also been formulated, to replace truth conditions with assertability conditions (Raatikainen 2004). There are certainly good textual reasons to believe that proofs play, at least, some role in Brouwer's conception of truth. For one thing, the notion of proof is employed in some definitions of concepts which we would not normally consider to be about proofs (i.e. not a concept belonging to proof-theory):

> Two mathematical entities are called different, if their equality has been *proved* to be absurd. (Brouwer 1952, 142, my emphasis)

A second reason is that the following two quotes are so alike that it is natural to interpret Brouwer as considering them noting but rhetorical variations of each other even though one has "true" where the other has "proved to be true", suggesting equivalence between them:

> Correctness of an assertion then has no other meaning than that its content has in fact appeared in the consciousness of the subject. We therefore distinguish between:
>
> 1. true
> 2. impossible now and ever
> 3. at present neither true nor impossible
>   *a.* either with, or
>   *b.* without the existence of a method which must lead to either 1. or to 2. (Brouwer 1951b)

> [I]n mathematics no truths could be recognized which had not been experienced, and that for a mathematical assertion a the two cases formerly exclusively admitted were replaced by the following four: 1. a has been *proved to be true*; 2. a has been *proved to be false, i.e. absurd*; 3. a has neither been proved to be true nor to be absurd, but an algorithm is known leading to a decision either that

> a is true or that a is absurd; 4. a has neither been proved to be true nor to be absurd, *nor do we know an algorithm leading to the statement either that a is true or that a is absurd.* (Brouwer 1955, 114, original emphasis)

It is clear that proofs are in some way constitutive of Brouwer-truth. Nevertheless, it cannot be correct to interpret him as *identifying* truth and existence of proof. For that is exactly what he (1954) forcefully criticizes the formalists for doing; they render mathematics meaningless by doing away with the content that is being proved. The Brouwerian would ask rhetorically: if there is nothing beyond the proofs, then what is it that is being proved? Trying to reduce truth to proofs is to put the cart before the horse; there has to be something more basic that proofs can be about. To prove something must be to show that it is true. If there is no independent notion of truth, then the concept of proof is taken as basic, and intuitionism is just a version of formalism.

Also, insofar as existence of proofs is admitted as partially constitutive of truth in Brouwer's view, it must be with a careful understanding of what proofs are. They cannot be understood as linguistic entities and they cannot even be something that is "build" from logic, for mathematics is independent of, and primary to, language and logic, according to an often repeated claim of Brouwer's.[5]

An attempt at a precision of the interpretation of Brouwer as equating truth with existence of proofs is what has become known as the "Brouwer-Heyting-Kolmogorov interpretation" or "BHK interpretation" for short. It gives the meaning of the logical connectives and quantifiers by recursively stipulating what counts as a proof of a sentence: a proof of $\phi \wedge \psi$ consists of a proof of ϕ and a proof of ψ (and the conclusion); a proof of $\phi \vee \psi$ consists of a proof of ϕ or a proof of ψ; a proof of $\phi \rightarrow \psi$ consists of a method for converting any proof of ϕ into a proof of ψ; a proof of $\neg\phi$ consists of a method for converting any proof of ϕ into a proof of a contradiction; a proof of $\exists x \phi(x)$ consists of an object d, a proof that d is in the given domain and a proof of $\phi(d)$; and a proof of $\forall x \phi(x)$ consists of a method for converting any object d in the domain into a proof of $\phi(d)$.

(To this story must be added an account of what a proof of an atomic sentence is. Such accounts are specific to the mathematical theory under consideration. Arithmetic can be formalized in such a way that the only atomic sentences are numerical equations, and then a proof of such a sentence of the form $a = b$ can be specified as something that begins with identity statements of the form $a = a$.)

[5]See, e.g., (1907, chapter 3), (1947) and (1952).

There are two specific problems for the BHK interpretation in addition to the already mentioned more general problems of the truth=proof interpretation. The first is that the interpretation of the disjunction is not faithful to Brouwer. He has it that *tertium non datur* already holds[6] for a proposition ϕ when the subject has a decision procedure for it, also prior to executing that procedure and thereby obtaining a proof of one of the disjuncts of $\phi \vee \neg\phi$. But the BHK interpretation does not allow us to assert that disjunction without being in a position to assert one of the disjuncts.

The second is, as Dummett (2000, 269–270) points out, that the definition, as it stands, is impredicative, because of the clauses for the conditional and the universal quantifier: A proof of $\phi \to \psi$ is a certain operation on all possible proofs of ϕ. We have no guarantee that we have a full grasp on what counts as a proof of ϕ before we have a full grasp on what counts as a proof in general, but that is just what is being defined.

I think the problem can be presented most forcefully in the form of a trilemma. Either 1) it is fixed in advance of the recursion on the complex sentences what counts as a proof of an atomic sentence or 2) it is not.[7] The first case can be subdivided into a case 1a) where these prefixed proofs include some that contain, as lines in the proofs, complex sentences, and 1b) where they do not. In all three cases there are unacceptable consequences which can all be exemplified with a proof concluding with modus ponens, i.e. a proof where the antepenultimate line is ϕ, the penultimate is $\phi \to \psi$ and the final line is the atomic sentence ψ. In case 1a) this proof is valid independently of the BHK recursion, so the proof stripped of its last line is a proof of $\phi \to \psi$ independently of the BHK recursion, making it redundant. In case 1b) this cannot be a proof of ψ, so no atomic sentence can be proved by modus ponens, which is absurd. And in case 2) it is consistent with the BHK interpretation to stipulate that ψ can be proved from ϕ no matter what these sentences are, for we can just claim that the mentioned proof is a valid proof for ψ if we also claim that the method *to any proof of ϕ, add a line containing $\phi \to \psi$ and then a line containing ψ* is a proof of $\phi \to \psi$, for that is then the method for converting any proof of ϕ into a proof of ψ required by the BHK interpretation of the conditional.

Dummett (1978a; 2000) has tried to improve on this situation by distinguishing between "canonical proofs" and a weaker notion of proof. Canonical proofs are some that never proceed via formulae that are more complex than the premises

[6]In his (1952) he writes that in this case "application of the principle of the excluded third is *permissible*"; in his (1908) that it is "*reliable* as a principle of reasoning" (my emphasis).
[7]The case of arithmetic belongs, as explained, in the former category, but it is not clear where other mathematical theories belong.

and conclusion. On the other hand, informal proofs or "demonstrations", the kind of proofs that you typically find in a mathematical paper, are some that, in principle, provide a method for obtaining a canonical proof. A disjunction may therefore be assertable by virtue of a demonstration, which, when converted into a canonical proof, would not only prove the disjunction but also one of the disjuncts. The BHK interpretation is then taken to define the weaker notion of proof, while presupposing only the notion of canonical proofs, and then the recursion is well-founded.

However, even if this works as a solution to the specific problems for the BHK interpretation, it does not resolve the more general problems of interpreting Brouwer as identifying truth and existence of proof. (But, it should be noted, it was not intended as an exegetical thesis by Dummett.) Just splitting the notion of proof into two different notions of proof cannot do that job. When I have nevertheless used the space to discuss it, it is because the idea has similarities to the exegetical thesis I will propose. For, according to this, Brouwer effectually splits the notion of truth into a strong and a weak variant.

The key, I believe, is to be found in one of the quotes above, namely the first one in this chapter: expected experiences are true only as anticipations; in their contents there is no truth. Here is a distinction between a strong notion of truth, truth-in-content (TIC), and a weaker, truth-as-anticipation (TAA).

The strong notion is what allows Brouwer to claim that "truth is only in reality", i.e. TIC is correspondence to an actual construction. Not only is an object, a, a construction, so is a property-ascription, $P(a)$, and a reduction to absurdity, $\neg P(a)$ (Brouwer 1907). To take a simple example (not Brouwer's own), a natural number n is a construction consisting of n elements; the property-ascription *n is the sum of two primes* is a one-to-one and onto mapping of these n elements to the elements of two primes; and the reduction to absurdity of the same property-ascription would consist in attempted constructions of such mappings between the n elements and all pairs of two primes smaller than n that are executed as far as possible until "the construction no longer *goes*" (p. 127). These mappings are themselves constructions, described by Brouwer (1908) as the predicate being "embedded" into the object.

If, on the other hand, the subject employs (intuitionistic) logic to deduce new truths from existing truths, the new truths are not necessarily TIC. It is just that the subject now knows how to make them true. He has an algorithm which will produce the truth-maker for the sentence and he knows in advance of executing it, that it will have that result. In other words, he can anticipate the TIC of the sentence; it is TAA:

> [T]here is a system of general rules called *logic* enabling the subject to deduce from systems of word complexes conveying truths, other word complexes generally conveying truths as well. [...] This does not mean that the additional word complexes in question convey truths *before* these truths have been experienced [...] (Brouwer 1948, 1243, original emphasis)[8]

Let us consider a few examples. The proposition that 17 is an odd number, is TIC when a mapping between 17 and two copies of 8 and a unit has been constructed. The proposition that $10^{10}+1$ is an odd number, however, is not TIC (except if the subject has been extremely industrious) but it is TAA if just the subject knows how to construct that number and a mapping between it and two copies of $5 \cdot 10^9$ and a unit. Further, assuming that our subject is not an expert on prime numbers, "$10^{10}+1$ is prime" is neither TIC nor TAA. He does have an algorithm for deciding the proposition (let us assume that he knows that much), but he does not know the result of executing it in advance. That brings us to the case of complex sentences: "$10^{10}+1$ is prime or $10^{10}+1$ is composite" is TAA, for the subject has an algorithm which he knows will make the sentence TIC. He just doesn't know how; he does not know which disjunct will become TIC, so neither of the disjuncts are TAA.

Based on this discussion, we can give the first part of a more precise, recursive formulation of when a sentence is TIC and TAA, respectively:

- $P(a)$ is TIC iff P has been embedded into a
- $\neg P(a)$ is TIC iff all options for embedding P into a have meet an obstacle
- $\phi \vee \psi$ is TIC iff ϕ is TIC or ψ is TIC (and the conclusion has been drawn)
- ◦ $P(a)$ is TAA iff an algorithm has been made which can make $P(a)$ TIC
- ◦ $\phi \vee \psi$ is TAA iff an algorithm has been made which can make either ϕ TAA or ψ TAA

Here "algorithm" means a method that not only is finite and would have the stated result if executed, but is known to the subject to be finite and to lead to that result. The notion of "embedding" is one that I will leave relatively vague as it is. Let me just make it clear that the idea is that TIC is an ontological rather than epistemological notion; when the subject has made a sentence TIC, he has not *verified* that it is true, he has *made* it true. (TAA on the other hand is connected with verifications.)

The clauses for conjunction are obvious, as are the generalizations of the clauses for the atomic sentences to predicates or arity more than 1:

[8] See also (Brouwer 1952, 141).

- $\phi \wedge \psi$ is TIC iff ϕ is TIC and ψ is TIC
- ∘ $\phi \wedge \psi$ is TAA iff an algorithm has been made which can make both ϕ and ψ TAA
- $P(a_1, \ldots, a_n)$ is TIC iff P has been embedded into $\langle a_1, \ldots, a_n \rangle$
- $\neg P(a_1, \ldots, a_n)$ is TIC iff all options for embedding P into $\langle a_1, \ldots, a_n \rangle$ have meet an obstacle
- ∘ $P(a_1, \ldots, a_n)$ is TAA iff an algorithm has been made which can make $P(a_1, \ldots, a_n)$ TIC

For existential quantification the clauses are analogous to those for disjunction:

- $\exists x \phi(x)$ is TIC iff $\phi(a)$ is TIC for some object a in the domain
- ∘ $\exists x \phi(x)$ is TAA iff an algorithm has been made which can construct an object a in the domain and make $\phi(a)$ TAA

The special BHK-interpretation of the universal quantifier and conditionals fits perfectly into the present interpretation on the side of TAA:

- ∘ $\forall x \phi(x)$ is TAA iff an algorithm has been made which can turn any object a for which it is TAA that a is in the domain, into the TAA of $\phi(a)$
- ∘ $\phi \to \psi$ is TAA iff an algorithm has been made which can turn TAA of ϕ into TAA of ψ

Truth-in-content is an extensional notion, and therefore the clause for TIC of a universally quantified sentence must be as follows:

- $\forall x \phi(x)$ is TIC iff all the objects a in the domain have been constructed and $\phi(a)$ is TIC for them all

This has the consequence that when the domain is infinite, a universally quantified sentence cannot be TIC, only TAA.

It is a more delicate matter what to say about TIC of conditionals. One thought would be that intuitionistic conditionals can only be understood in the algorithmic sense of the BHK-interpretation. In that case, TIC should never be attributed to conditionals. Insofar as it should, being an extensional notion, it seems most reasonable to understand it in terms of the classical definition of the conditional:

- $\phi \to \psi$ is TIC iff $\neg\phi$ is TIC or ψ is TIC

Adopting this clause implies accepting the inference from $\neg\phi$ to $\phi \to \psi$ for any ϕ and ψ. There is no indication in Brouwer's writings that he does; see (van Atten 2009) (where it also noted that Heyting, Kolmogorov, Troelstra, van Dalen, Martin-Löf and Dummett do accept it).

The TIC of $\neg\phi$ has only been defined for atomic ϕ so far, which must be remedied. But again, the extensionality of TIC, i.e. the requirement of correspondence with actual constructions, settles the issue unequivocally. A negation of a complex sentence being TIC must be understood in terms of the classical equivalence with a sentence where negations have narrower scope, so as to be reducible to the already defined TIC of atomic sentences and negations of atomic sentences:

- $\neg\neg\phi$ is TIC iff ϕ is TIC
- $\neg(\phi \vee \psi)$ is TIC iff $\neg\phi$ is TIC and $\neg\psi$ is TIC
- $\neg(\phi \wedge \psi)$ is TIC iff $\neg\phi$ is TIC or $\neg\psi$ is TIC
- $\neg(\phi \to \psi)$ is TIC iff ϕ is TIC and $\neg\psi$ is TIC
- $\neg(\exists x \phi x)$ is TIC iff all the objects a in the domain has been constructed and $\neg\phi a$ is TIC for them all
- $\neg(\forall x \phi x)$ is TIC iff $\neg\phi a$ is TIC for some object a in the domain

That just leaves us with TAA of negated sentences to be defined:

- ○ $\neg\phi$ is TAA iff an algorithm has been made which can turn TAA of ϕ into some construction and the obstruction of the same construction

This interpretation of Brouwer is, in a sense, a combination of the three proposed interpretations I discussed above: truth as actual construction, truth as potential construction and truth as proof. Both TIC and TAA consist of actual constructions. TIC consists in the actual construction of "things" and "qualities of things" in the language of the quote at the beginning of this chapter, and TAA consists in the actual construction of "rules". As such, both notions of truth are tensed. On the other hand, the admittance of the weaker[9] notion of truth, TAA, is due to a reliance on potential constructions. It is a trust in the possibility of, to a certain extend, predicting the properties of not yet effected constructions which justifies anticipated-truth when there is not yet truth in the strong ontic sense. And finally, TAA is identified with the existence of proof. But it is in a sense of proof where it does not necessarily

[9]For any sentence ϕ, ϕ being TIC implies that ϕ is TAA. The TAA-making algorithm is the "empty" algorithm that is vacuously executed.

have to be a linguistic entity. Rather, an intuitionistic proof is a method for producing the truth-maker of the given sentence (although the subject needs to know that the method does that, and that knowledge may be the result of a proof in the traditional sense of the word). In this way, TAA is grounded in TIC. TAA of complex sentences is defined in terms of TAA of simpler sentences, and TAA of atomic sentences is defined in terms of TIC. Thus, there is no problem of impredicativity as in the BHK interpretation.

That a sentence is TAA means that it would become TIC if the algorithm in question were executed along with the algorithms, corresponding to simpler sentences, thereby produced and the... etc. etc. down to atomic sentences.[10] However, when there are universal generalizations over infinite domains or negated existential claims over same involved, that task is impossible, as it consists in the execution of an infinity of algorithms. Still, the grounding of TAA in TIC is not thereby nullified, for any given finite part of the infinite conjunction, which such a sentence amounts to, can be realized as TIC.

The present interpretation accommodates both the 1951b and the 1955 quotations above, without having to resort to the highly problematic "reduction" of truth to existence of proof. The four possibilities for the status of an assertion a in the former quote are 1) that a is TIC, 2) that $\neg a$ is TIC, 3a) that neither a nor $\neg a$ is TIC but $a \vee \neg a$ is TAA, and 3b) that neither a nor $\neg a$ is TIC and $a \vee \neg a$ is not TAA. In the latter quote, the four possibilities distinguished are 1) that a is TAA, 2) that $\neg a$ is TAA, 3) that neither a nor $\neg a$ is TAA but $a \vee \neg a$ is TAA, and 4) that neither a nor $\neg a$ nor $a \vee \neg a$ is TAA. These are different categorizations but they both give four possibilities which are mutually exclusive and collectively exhaustive.

With this interpretation we can also explain how Brouwer can deny in the π example that it was true at his time that a k_1 exists without denying the predetermination of the sequence of decimals of π. For any claim about the value of a specific decimal, the subject can anticipate finding the answer, in the sense that he knows that he will find it within a preknown number of construction steps if he goes through the appropriate procedure. If he starts looking for a k_1 by searching through the decimals one by one, he can not anticipate finding one, he can merely hope for it.

Even though the issue has already been touched upon, let me explicate the consequences for the semantics of disjunctions. The case of TIC is simple: if a

[10] In the case of universally quantified sentences, conditionals and negations there are prerequisites for doing so: for universally quantified sentences one would need to have constructed the entire domain; for conditionals (assuming that the above clause for TIC of such is adopted) one would need the TAA of the antecedent; and for negations one would need, *per impossibile*, the TAA of the negated sentence.

disjunction is TIC then at least one of the disjuncts is too. But a disjunction can be TAA without any of the disjuncts being so. In particular $\tau \vee \neg\tau$ is TAA if the subject has a decision procedure for τ, but if he does not know in advance on which side the procedure will come out, neither τ nor $\neg\tau$ will be TAA. This allows Brouwer to state that

> Each assertion τ of the possibility of a construction of bounded finite character in a finite mathematical system furnishes a case of realization of the principle of the excluded third (Brouwer 1948, 1245)

For in a finite system the procedure "try all possibilities" is a decision procedure; it will result in either τ or $\neg\tau$ becoming TIC and *a fortiori* TAA. So TAA does not distribute over disjunctions, but it is, so to speak, distribut*able* over disjunctions with a little work. That is, a disjunction being TAA at a given point in time does not imply that either of the disjuncts is TAA at that time, only that one of the disjuncts can be made TAA by executing the algorithm that makes the disjunction TAA.

We are not in possession of a decision procedure for Goldbach's Conjecture, $\forall nP(n)$, and hence $\forall nP(n) \vee \neg\forall nP(n)$ is not TAA. But for each n, we do have such a procedure for $P(n)$, making $P(n) \vee \neg P(n)$ TAA. So the algorithm consisting of "plugging" the given n into that procedure, is the algorithm which turns any object in the domain $\mathbb{N}$ into the TAA of $P(n) \vee \neg P(n)$, required for $\forall n(P(n) \vee \neg P(n))$ being TAA.

It may also be worth explicitating why the clause for TAA of a disjunction does not read "$\phi \vee \psi$ is TAA iff an algorithm has been made which can make either ϕ TIC or ψ TIC". Let again τ be a decidable but undecided proposition. Then this proposition, where n ranges over the natural numbers, is TAA (for anyone in possession of the decision procedure and aware of the following): $\forall n(\tau \wedge n = n) \vee \forall n(\neg\tau \wedge n = n)$. By deciding τ, one of the disjuncts becomes TAA but not TIC, for the latter would presuppose a completed construction of all the natural numbers.

3.1 An equivalence in propositional logic

In this section, and the following two sections, I will confront the interpretation with three examples of Brouwerian mathematics to show how it can account for them. The three examples are those that van Atten in his (2012) brings forward in defense of the claim that the "B" does rightly belong in the name "BHK-interpretation", and I have copied his subsection headings.

The first is Brouwer's proof of the logical law $\neg\neg\neg A \leftrightarrow \neg A$. This proof van Atten uses (sections 2.2 and 3.1.1) to argue against an understanding of conditionals $A \to B$ as meaning just "$A \wedge B$ with the extra information that the construction for B was obtained from that for A" (which would be a consequence of the truth-as-actual-construction interpretation[11]):

> The argument begins by pointing out that $A \to B$ implies that $\neg B \to \neg A$ [...] It would not have been possible for Brouwer to make this inference if at the time it would have been among his proof conditions of an implication to have a proof of the antecedent, as then a proof of $A \to B$ would lead to a proof of B and thereby make it impossible to begin establishing the second implication by proving its antecedent $\neg B$.

Obviously, this is not an argument specifically for the BHK-interpretation, only against the mentioned alternative. Its conclusion is also consistent with the present interpretation, where $\neg\neg\neg A \leftrightarrow \neg A$ is TAA. Or rather: $\neg\neg\neg A \leftrightarrow \neg A$ is TAA for anyone who has understood the following proof (or one like it), as it provides a method of turning the TAA of $\neg\neg\neg A$ into the TAA of $\neg A$ and *vice versa*:

We first prove the TAA of $A \to \neg\neg A$. That is done by providing a method of turning TAA of A into TAA of $\neg\neg A$. So assume that A is TAA. TAA of $\neg\neg A$ is an algorithm for turning TAA of $\neg A$ into a construction and the obstruction of the same construction. So assume also that $\neg A$ is TAA. Use that to turn the TAA of A into a construction and the obstruction of the same construction. Discarding the second assumption, we have the TAA of $\neg\neg A$. And by also discarding the first assumption, the TAA of $A \to \neg\neg A$ is reached.

Second, we prove that the TAA of $A \to B$ implies the TAA of $\neg B \to \neg A$. Assume the antecedent and the TAA of $\neg B$. The following is an algorithm for turning the TAA of A into a construction and the obstruction of the same construction, i.e. the TAA of $\neg A$: use the first assumption to turn the TAA of A into the TAA of B and then use the second assumption to turn that into a construction and the obstruction of the same construction.

A special case of the first proposition proved is that $\neg A \to \neg\neg\neg A$ is TAA. And the two propositions together imply the TAA of $\neg\neg\neg A \to \neg A$. The algorithm which makes $\neg\neg\neg A \leftrightarrow \neg A$ TAA is then simply the concatenation of these two algorithms.

This example is one that the BHK-interpretation and the "two truths" interpretation can account for equally well. I will argue that the next two are some where the latter does better than the former.

[11] And a possible interpretation of the TIC of $A \to B$, instead of the one above.

In this section I have considered a single logical theorem. But, the more general claim that this interpretation is complete with respect to intuitionistic first order predicate logic is also true. It is true in the following sense: for each axiom and inference rule of that logic there is a certain (quite simple) proof, such that the axiom/inference rule is TAA for anyone who has understood the proof.

3.2 The proof of the Bar Theorem

The next example considered by van Atten is the proof of the Bar Theorem (Brouwer 1924a; 1927; 1954; 1981), which can be stated thus: *if B is a decidable bar on a spread, then B contains a well-ordered thin bar.* I will begin this section by explaining the terms used in this formulation, before interpreting Brouwer's proof of, and his comments on, it in the light of the two truths. As this will get somewhat abstract and perhaps difficult to follow, the section will conclude with a toy example to make matters more concrete.

For present purposes we can define a spread as a species of tuples of natural number, in that we consider the empty tuple as a such, which satisfies the following. First, it is decidable for any tuple whether or not it is in the spread. Second, if $\langle a_1, \ldots, a_n \rangle$ is in the spread then so is $\langle a_1, \ldots, a_{n-1} \rangle$. Third, if $\langle a_1, \ldots, a_n \rangle$ is in the spread then there exists (in the intuitionistic sense of the word) a natural number a_{n+1} such that $\langle a_1, \ldots, a_n, a_{n+1} \rangle$ is also in the spread.

A tuple $\langle a_1, \ldots, a_n \rangle$ is called an ascendant of a tuple $\langle a_1, \ldots, a_n, \ldots, a_{n+m} \rangle$ in the spread, and the latter is called a descendant of the former. If $m = 1$ the modifier "immediate" may be added.

It is helpful to mentally picture a spread as a tree in which each tuple is a node with all its immediate descendants as nodes immediately below it. An infinite route from the root, i.e. the empty tuple, downwards then corresponds to a choice sequence.[12]

A bar is a subspecies of a spread such that every choice sequence "in" the spread has an initial segment (one of the nodes it goes through) in that subspecies. A bar can be pictured as an area stretching the entire breadth of the tree so that every choice sequence must pass through it. It was discovered by Kleene and Vesley (1965) that for the theorem to hold it must be assumed that the bar is decidable, which is to say that it is decidable whether a given tuple is in it or not.

[12]At this point the reader can temporarily ignore the critique of choice sequences in the previous chapter, for the infinitude of the sequence is not really relevant here.

For the purpose of being a bar, such an area does not need to be deeper than one node; hence the notion of a thin bar which is a bar such that for any tuple in it, no ascendant or descendant of it is also in the bar.

Well-orderings are defined inductively. A one-element species is a well-ordering, and if $A_0, \ldots, A_n$ or $A_0, \ldots$ are disjoint well-orderings, then their union, equipped with the following ordering, is also a well-ordering: $x < y$ if either x is from an A_i and y is from an A_j such that $i < j$ or x and y are from the same A_i and ordered $x < y$ therein. The A_i's are called "constructional subspecies" of the resulting well-orderings.

For this analysis, there is one specific detail about the ontology of well-orderings that is important. Brouwer (1981, 44) demands that the A_i's are in the "available stock" of already constructed well-orderings, before they can be used to construct a larger well-ordering. This may suggest a demand for strict "bottom-up" construction, which is misleading. First of all, one should of course remember that if a well-ordering is constructed out of infinitely many other well-orderings, then the infinity is potential, so they cannot all have been previously constructed. It follows that the existence of these well-orderings must be understood as true-as-anticipation.

In other words, Brouwer accepts that the larger-scale structure is constructed prior to the smaller-scale details of that structure. This conclusion is reinforced by an example from (Brouwer 1981, 49), where a well-ordering is constructed with the aid of a fleeing property (see Section 1.5). Pretend that such a property is given and let k be the (hypothetical) least natural number with that property. Further, let A_i be an ω-sequence for $i < k$ but just a one-element species for $i \geq k$. Brouwer takes these A_i's to be acceptable building blocks for a well-ordering. So, not only can the building blocks be constructed after the house, we can also be largely ignorant about the shape and size of these building blocks.

This brings us to the end of the explanation of what the theorem says. We can turn to the interpretation of its proof.

The Bar Theorem has the form of an implication, and the proof turns on considerations of how possible proofs of the antecedent can be manipulated into a proof of the consequent. This fits with the BHK-interpretation. But, what Brouwer means by "proof" in that context is very different from the normal understanding of the word.[13] When his non-standard use of the word is taken into account, the two truths interpretation can explain Brouwer's proof

[13] See (Sundholm and van Atten 2008) regarding Brouwer's use of the words "proof", "demonstration", and "argument" and their equivalents in German and Dutch.

in a way that is more detailed and less prone to misunderstandings than the BHK-interpretation.

Brouwer writes that any proof of the antecedent can be expanded into a so-called canonical proof. Such a canonical proof is an infinite (if the part of the tree "above" the bar is) mental construction with a structure that is itself a well-ordering. When the word "proof" is taken in the normal sense, that is a baffling claim.[14] How can something persuade us of the truth of a proposition if it is infinite and therefore unsurveyable? And how can Brouwer be certain that a proof must be expandable into the form he describes? The word "proof" (or "demonstration") must be understood differently:

> Intuitionistically, to give a demonstration of a mathematical theorem is not to produce a certain linguistic object, but to produce a mental mathematical construction (or a method to obtain one, which method is of course also a mental mathematical construction) that makes the corresponding proposition true. Therefore, the requirement, for a demonstration that the consequence
>
> $$\text{A is true} \Rightarrow \text{B is true}$$
>
> holds, of a method that transforms any demonstration that A is true into one that B is true, is really the requirement of a method that transforms any mathematical construction that makes A true into one that makes B true. (Sundholm and van Atten 2008, 61)

This is correct, but also easy to misunderstand: is a "mathematical construction that makes A true" not an infinite structure which can never be completed and thereby ready for being transformed into a mathematical construction that makes B true? With the present interpretation, with its distinction between two ways that A and B can be true, we can make this more precise and avoid the confusion. A proof of the implication is a method for transforming the TAA of A into the TAA of B, not a method to transform the actually infinite structure that would make A TIC into the actually infinite structure that would make B TIC.

Getting to the specific details of the proof of this theorem, the antecedent says, when interpreted in the appropriate intuitionistic way, that when I construct a choice sequence in the spread, I will at some point construct an initial segment that is in the bar; that I will be able to determine that the initial segment is in the bar, when it is; and that I know in advance that it will happen within a

[14] For example, relying on a truth=proof interpretation of intuitionism, Epple (2000) comes to the conclusion that the proof of the Bar Theorem does not live up to Brouwer's own epistemological standards.

calculable number of construction steps.[15] The antecedent is of the ∀∃-form; for all choice sequences in the spread, there exists an initial segment of it that is in the bar. So the antecedent being TAA means that for any given choice sequence, an initial segment of it that is in the bar can be constructed, i.e. that the instance of the universal generalization for the specific choice sequence can be made TIC. It is hopeless to expect that we could know how any possible proof, in the normal sense of that word, that could make that antecedent TAA would look like, but it is trivial to see how this TIC is accomplished: just construct the terms of the choice sequence one after another, and each time run the decision procedure for bar-membership on the resulting initial segment. The infinite mental construction, that Brouwer misleadingly calls a canonical proof, is the (unaccomplishable) TIC of all the instances, i.e. of the antecedent.

Therefore, what Brouwer means by "canonical proof" is very different from how Dummett understands the term. The reasoning that makes the subject know in advance of constructing a choice sequence, that it will meet the bar, is not the thing that can be transformed into a canonical form. Rather, the canonical form is the imagined infinite result of applying an ability to construct a certain kind of objects (choice sequences hitting bars) in all possible ways, and that ability can be transformed into an ability to construct another kind of objects (well-orderings).

Let me elaborate in a way that brings us closer to Brouwer's own formulations. He calls elements of the bar "secured" nodes, while nodes above are "unsecured" but are "securable" when it is established that all choice sequences through it hit a secured node. The securability of a node is accomplished through induction from below; if all the immediate descendant of a given node are securable, then that node is securable. That is the induction step and is called an "elementary inference". The securability of nodes high up in the tree, in particular the root, is reached through repeating such steps; if the tree is large enough, infinitely many of them. In his formulations Brouwer indulges in the fantasy that we could actually go through the entire construction process, bottom-up, but of course he should not be interpreted literally; it is a mere *façon de parler*. Actually, we can just construct finitely many nodes, top-down, and must have some means, independently of actually constructing *all* the nodes that contributes to their securability, of knowing that all choice

[15] The last bit is not precise: "the algorithm in question may indicate the calculation of a maximal order n_1 at which will appear a finite method of calculation of a further maximal order n_2 at which will appear a finite method of calculation of a further maximal order n_3 at which will appear a finite method of calculation of a further maximal order n_4 at which the postulated node of intersection must have been passed. And much higher degrees of complication are thinkable." (Brouwer 1954, 13)

sequences through them will hit the bar. Such means are of course arguments or, in the normal meaning of the word, proofs.

The strong similarities between constructing well-orderings and establishing securability of nodes should now be clear. First of all, obviously, the elements of the thin bar correspond to the one-element well-orderings, and the nodes above it to the well-orderings that are constructed from smaller well-orderings. In addition, for both securability and well-orderings, the literal reading of Brouwer suggests a strict bottom-up constructional process in which a brick in the wall can only be added, when all the bricks it rests on are in place; but instead it must be seen as a potentially infinite top-down process where the status of being securable/being a well-ordering steams from prior knowledge that whatever choice sequence/series of constructional subspecies of constructional subspecies of constructional subspecies etc. the subject may *actually* construct, it will hit the bar/bottom out in a one-element well-ordering.

The ability that the subject must posses (and know to posses) in order for the antecedent to be TAA is virtually the same as the ability that makes the consequent TAA. Hence, proving the Bar Theorem is actually trivial. To prove the TAA of the implication, what is needed is a way of turning any method for constructing any given "part" of the truth-maker ("TIC-maker") of the antecedent into a method for constructing any given part of the truth-maker of the consequent. But doing the former is essentially the same as doing the latter, so any method for the former is almost a method for the latter, making the "turning" of the one into the other trivial.

Brouwer himself notes this triviality in (1927, original text: 63fn7, English translation: 460fn7). That is a comment that can be explained with the two truths interpretation, and this explanatory success is, I believe, a point in its favor.[16]

Now to the promised toy example. In Figure 3.1 is the top part of a spread where the natural numbers allowed are restricted to 1 and 2. It is equipped with a very simple five element bar which contains a well-ordered three element thin bar. According to the bottom-up story, the subject should construct the well-ordered thin bar by first constructing $\langle 1,1\rangle$,[17] $\langle 1,2\rangle$ and $\langle 2\rangle$, then constructing $(\langle 1,1\rangle,\langle 1,2\rangle)$ and finishing with $((\langle 1,1\rangle,\langle 1,2\rangle),\langle 2\rangle)$ (the brackets indicate constructional subspecies). Of course that is quite feasible in this

[16] The above discussion could perhaps have been improved in precision with a formalization of the Bar Theorem. But it is not clear to me that such a formalization is possible. An attempt at a formalization can be found in (Kleene and Vesley 1965, 52), but there the consequent is rendered as a principle of backward induction, which I think is unfaithful to Brouwer.

[17] I omit "intention to expand".

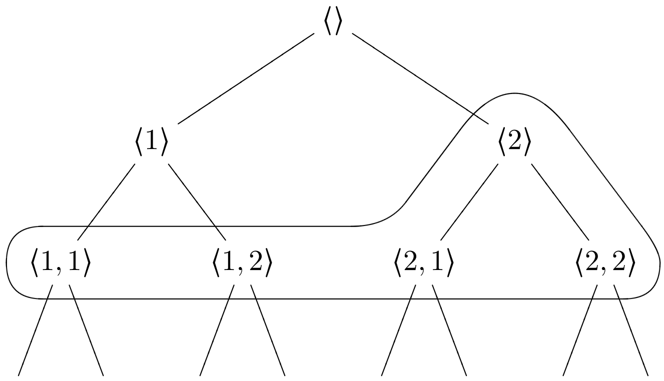


Figure 3.1

case, the thin bar being finite. But for the purpose of the toy example, let us pretend that the subject (for whatever mundane reason) only has time to do two construction steps, and let that restriction simulate finitude. Then the existence of the bar and the existence of the well-ordered thin bar cannot become TIC.

But he can construct any part of the bar, top-down. For instance the first construction step could result in $\{\langle 1\rangle^\star, \dots\}$ and the second in $\{\langle 1, 2\rangle, \dots\}$ (the curly brakets indicate a species (as a bar is) with some intensional criterion of membership not displayed; the dots indicate that it is incomplete qua extensional object; and the star indicates that $\langle 1\rangle$ is not itself an element of the bar but has to be further developed). The ability to "fill out" any part of this species and the knowledge that any "starred element" can be developed into an element of the species no matter which natural number (here 1 or 2) is chosen in the following steps is what constitutes the TAA of the antecedent of the Bar Theorem.

That ability and knowledge is, as noted, virtually the same as the the ability and knowledge that constitutes the TAA of the consequent. Here the corresponding top-down construction of a part of the well-ordering has as its first step $(\langle 1\rangle^\star, \dots)$ and as its second $((\dots, \langle 1, 2\rangle), \dots)$. The difference is simply that in the construction of (parts of) the well-ordered thin bar, some extra structure from the construction process is preserved.

3.3 Ordering axioms

The last example (van Atten 2012, section 3.1.3) is concerned with Brouwer's definition of so-called "virtual orderings". These are given through five axioms, of which one, serving as example, will be sufficient for present purposes, so let us take the simplest, number five: "From $r < s$ and $s < t$ follows $r < t$".

The following comment of Brouwer's, concerning these axioms, is cited by van Atten as a confirmation of the BHK-interpretation:

> The axioms II through V are to be understood in the constructive sense: if the premises of the axiom are satisfied, the virtually ordered set should provide a construction for the order condition in the conclusion.

Van Atten claims that "[t]his is a clear instance of the clause for implication in the Proof Interpretation". But the BHK-interpretation renders the axiom as "any proof of $r < s$ and $s < t$ must be convertible into a proof of $r < t$". However, the operative words in Brouwer's comment are "satisfied" and "provide a construction" which are more specific than the ambiguous "proof". The "two truths" interpretation captures this comment much better. For the TAA of the axiom, i.e.

TAA of (from $r < s$ and $s < t$ follows $r < t$)

is equivalent to

TAA of ($r < s$ and $s < t$) can be turned into TAA of $r < t$,

which is the same as

TAA of ($r < s$ and $s < t$) can be turned into an algorithm which makes $r < t$ TIC.

This seems to be a much more reasonable explicitation of the comment.

3.4 Two truths versus BHK

In the original 6th century Indian version of chess, the winning criterion was to actually capture the opponent's king. Only later did the Persians amend the rules so that a player would win already when the king was made check mate (Davidson 1949). The original version is the most intuitive and it would be difficult to imagine that the game could have been invented directly in the Persian form; the concept of check mate is difficult to explain except when done in terms of what *would happen* in the next round of the game. On the other hand, the Persian "contraction" of the game makes good sense, as that final round is trivial and not worth actually executing.

This makes for a nice analogy: TIC is like actually capturing the king, while TAA corresponds to making the king check mate or, more generally, being in possession of a winning strategy for the Indian version of the game. If the existence of such a winning strategy is common knowledge to the players, then there is no point in actually completing the game – but only by reference to the possibility of actually capturing the king does the winning strategy make sense.

My objection to Brouwer and his interpreters is that they have not made this distinction clearly. In chess, the "contraction" of the game is so simple that anyone presented to the Persian version can easily see the connection with the Indian version, and therefore think like an Indian while acting like a Persian. The "contraction" of TIC to TAA is quite complex. Therefore any introduction to intuitionism should begin by clearly explaining TIC and only then move on to the less basic and more abstract concept of TAA, which is what the proofs in intuitionistic papers make contact with.

The clauses of the BHK-interpretation are especially objectionable in their conflation of the two kinds of truth: the clause for disjunction is only correct for TIC; while the clauses for the conditional and the universal quantifier are only correct for TAA. I think the BHK-interpretation is comparable to early analysis in that only experts can interpret the interpretation the right way. Anyone learning about it for the first time is almost bound to get it wrong.

There is an obvious problem for my interpretation and the critique of BHK that needs to be considered, namely that Brouwer explicitly endorsed Heyting's interpretation:

> [W]hile preparing a note on intuitionism for the Bulletin of the Royal Academy of Belgium, I was pleasantly surprised to see the publication of a note of my student Mr. Heyting which elucidates in a magisterial manner the points that I wanted to shed light upon myself. I believe that after Heyting's note little remains to be said. (van Dalen 2013, 607)

There are a couple of reasons why I do not attribute much weight to this endorsement. First, given that formalism is not something Brouwer cares for, he could easily have made that remark without really having thought it through. And even if he did, it is a commonplace that you find out that a given formulation of your position, that you first thought to be perfectly precise, turns out to be improvable. Also, Brouwer's stamp of approval on Heyting's clauses was within the context of a discussion about whether intuitionism introduces a third truth value. Thus, the approval can be mostly due to the fact that Heyting got that part right.

3.5 Lawless choice sequences revisited

By qualifying the notion of "algorithm", we can extend the two truths interpretation to also cover choice sequences that are not lawlike. Let us return to the example of the choice sequence α, of which only the first three elements have been chosen and the sole restriction on future choices is that they must be natural numbers. Consider the following example sentences:

1. The 17th term of α is 99.
2. The 17th term of α is not 99.
3. The 17th term of α is 99 or the 17th term of α is not 99.

Neither of these sentences is TIC. Brouwer would say that the last sentence is true (Brouwer 1908), while the first two sentences are not. With the right understanding of "algorithm", this fits with the given clauses for TAA.

With an understanding of the word that is too rigid, sentence 3 would come out as not TAA. The procedure that makes the sentence TIC is the one described by the instruction "choose additional 14 elements of α". On a narrow understanding this is not an algorithm because it involves choices.

On the other hand, we cannot replace "algorithm" with something as broad as "method", for that would over generate sentences that are TAA: the subject has a method for making sentence 1 TIC, namely deciding to pick 99 as the 17th element.

The right understanding is most clearly explained with a story about *two* subjects. One subject is the generator of α and chooses one new element thereof, whenever he is prompted to do so by the second subject. The second subject is the one for which sentence 1 and 2 are not TAA while sentence 3 is. She has an algorithm, in the strict sense of the word, that will make sentence 3 TIC (by making one of the disjuncts TIC), namely simply *14 times in a row, ask the first subject for a new element of α.* Hence, sentence 3 is TAA for her, while sentences 1 and 2 are not, as she has no influence on what numbers are chosen.

Having to refer to two different subjects is not in the spirit of Brouwer who emphasizes the individual. It can be avoided if we imagine a subject who manages to keep his tasks separate – that is, when he chooses elements of a choice sequence, he chooses freely within the explicit restrictions he has imposed on himself without being influenced by the judgments he himself has previously made about that same choice sequence.

We have thus reached a more precise formulation of the intuitionist stipulation discussed in Section 2.5. The added precision does not, however, substantially affect the force of the critique that this stipulation about the semantics for lawless sequences is arbitrary. It does not, because we can similarly criticize the more general adoption of TAA as the most extensive notion of truth for being unnecessarily restrictive. That is the agenda for the next chapter.

Chapter 4

Non-verificationist constructivism

So, according to Brouwer there are these two kinds of truth and no others. What has actually been constructed makes for truth makers for a strong kind of truth. And when the creating subject has foreknowledge about how future constructions will go, that can, in a limited way, provide for a weaker kind of truth. But no even weaker form of truth exists. There is no sense in which it is true that 2 plus 2 equaled 4 prior to the birth of the first human being. That is his highly counterintuitive claim.

I hope that with the attempt in the previous chapter to make it more explicit than it has been previously what Brouwer's thesis is, it is possible to make it clear, not only what is so counterintuitive about the claim, but also why it does not follow from the constructivist ontology. That, and formulating an alternative that gets us closer to classical mathematics, cf. the overall goal as described in the introduction, is the aim of this chapter.

A possible alternative position would be that there are only truth-in-content and nothing beyond it. That is the conclusion one will be compelled to come to if one thinks any true mathematical sentence must have a truth maker and such a truth maker can only consist of the actual existence of the structure described in the sentence and, furthermore, that such structures can only be provided by the mind. Brouwer does not claim that. He believes that truthful things can be said about future constructions.

Another possible alternative position is that everything about lawlike future constructions is true in advance. This is the thesis that I will defend below, and there I will formulate it more precisely. For now, an example should suffice: the Goldbach Conjecture is true or false at present *because* it is determined in advance for each even number what would happen if one were to construct it and test whether it is the sum of two primes.

Brouwer holds an intermediate position, according to which future constructions can do the job of truth makers in some cases but not in all. The big question is *why*. It is not difficult to see what may attract someone to one of the two alternative positions, but it is not immediately obvious why one would think that the criteria for being a truth-as-anticipation are exactly those that mark the limit of truth. Why is it that *tertium non datur* holds for assertions about a huge not-yet-constructed structure just because it is finite, while it does not for Goldbach's Conjecture even as late as the second before someone were to find a counter-example?

Brouwer himself does not provide much of a reason. His philosophical arguments are mostly negative, directed at the Platonist. His official alternative platform is that of extreme ontological austerity, which points in the direction of the TIC-only position. His departure from that position seems to be due to a strong objectivist intuition. However, arguments are lacking in order to explain why we should only follow the objectivist intuition some of the way and not continue ahead to the other alternative position mentioned above.

Hence, we must look elsewhere. The only developed explicit argument for this intermediate position is Dummett's, so the first section of this chapter is devoted to him. However, Wittgenstein's rule following considerations may also seem to offer the hope of a justification. So to him I will turn in Section 4.3. Finding these arguments unsuccessful, I develop the alternative position in Sections 4.2 and 4.4 and the following chapters. The aim is to drive a wedge between constructivism and verificationism and show that, if nothing else, the project of developing a non-verificationist constructivism is at least as legitimate as the Platonist and the intuitionist projects – although, of course, that is to understate the ambition.

4.1 Dummett's argument from meaning constraints

Very much contrary to the spirit of Brouwer's philosophy, Dummett's argument in defense of Brouwer takes for its point of departure the social character of language: as meaning must be shared between the members of the language community, the meaning of a sentence cannot go beyond what it is possible to learn for a member of that community. Dummett tries to convince us that the classical semantics for mathematics cannot be learned and is therefore no semantics at all but just a mere illusion of understanding among those who practise classical mathematics.[1]

Dummett contrasts two overall types of theories of meaning for natural language. The first is the kind of theory according to which the meaning of sentences is given by their truth conditions, as suggested by Frege (1884), Russell (1919) and the early Wittgenstein (1921).[2] We will follow Dummett in calling this "truth-conditional semantics" (even though this is somewhat misleading as we have just interpreted Brouwer as understanding sentences in terms of not just one but two sets of conditions for when a sentence is true, and because, according to Dummett, *any* semantics can be formulated in terms of what it takes for a sentence to be true, if we understand that broadly enough). This kind of semantics is characterized by having truth-conditions that are independent of what is and can be known. The second is the kind of theory that defines meaning via conditions of justification. Dummett's argument for intuitionism consists of two sub-arguments. The conclusion of the first is that truth-conditional semantics has the absurd consequence that language cannot be learned at all. The conclusion of the second is that justificationist theories of meaning cannot deliver classical semantics for mathematics but only intuitionist semantics.

The sub-argument for the unlearnability of a language with truth-conditional semantics is a quite simple one from two premises. One premise is that the relation of dependency among sentences learned by an individual has to be well-founded: someone can learn the meaning of a new sentence by having it explained with other sentences, but of course, those sentences have to be antecedently grasped by the individual, so learning language in childhood cannot commence with linguistic explanations. The second premise is that truth-conditions are given in the form of instances of the T-schema:

$$(\mathrm{T_P}) \qquad \text{“P” is true iff P}$$

Therefore, the argument goes, understanding the sentence P presupposes understanding the sentence[3] $\mathrm{T_P}$, so, by generalization, every indicative sentence can, according to the truth-conditional theory of meaning, only be understood if some other sentence is already understood, which by the first premise leads

[1]This section is based on Dummett's (1978b), (1991a), (1991b), (1993), (2000) and (2006). Of central importance are chapters 4 and 5 of (2006), the paper "The Philosophical Basis of Intuitionist Logic" in (1978b), chapters 14 and 15 of (1991b) and the first paper in (1993).

[2]For an historical overview, starting with these thinkers and covering the development up to the present, see (Stanley 2008).

[3]Dummett (2006) holds to a "language before thought" doctrine, but claims that a similar argument goes through on a "thought before language" doctrine as well.

to the absurdity that no one understands any indicative sentence.[4] In contrast, according to Dummett, justificationist theories of meaning are in line with the demand for well-foundedness, because they give the meaning of sentences, not in terms of a theoretical knowledge like the truth-conditional theories, but as a practical ability to decide when asserting the sentence has been justified.

However, the conclusion of the second sub-argument states, a justificationist basis can only deliver intuitionistic logic for mathematics. For a universal quantification, to take that example, over an infinite domain would only satisfy the law of bivalence, according to the justificational meaning-theories, if there was a method that would always deliver a verification or a falsification when executed. Obviously, there *is* no such method, but Dummett considers an appeal that his opponent might make to counterfactual circumstances where there is one: The classical logician may bring in supertasks in her defense, if she concedes that the logical rules have to be warranted on a justificationist foundation. She could do that by claiming that what it means for a universally quantified sentence to be true is that if an agent with super-human powers were to go through all its instances, he would verify each one, and conversely what it means for it to be false is that if that agent worked his way through all the instances, he would falsify at least one. (Set aside the issue that we are then limited to cases where the agent can enumerate the instances and where each instance is decidable.)

Dummett counters this defense of classical logic with a general point about subjunctive conditionals, namely that if they are true or false, they cannot be so in a basic, or as Dummett puts it "barely", way; they have to be true/false in virtue of the truth/falsity of some other, categorical statement. The classical logician is here identifying the truth of the classical universal quantification $\forall x \phi(x)$ with the truth of a subjunctive conditional of the form $S \rightarrow \forall x \phi(x)$, where the antecedent is a statement to the effect that the supertask has been carried out, and the consequent has to be interpreted as, in our terminology, truth-in-content. She also identifies the falsity thereof with the truth of

[4]The most explicit formulation of this argument can be found in (Dummett 2006), where he first writes that the "truth-conditional account of sense makes a grasp of sense unequivocally into the possession of a piece of theoretical *knowledge*" (p. 48, original emphasis) and then continues "if we attempt to explain the understanding of a sentence as consisting in the possession of a piece of knowledge about that sentence, our explanation is circular: we are trying to explain grasping one proposition – that asserted by the sentence – in terms of judging another – the proposition that the sentence is true under such-and-such conditions – to be true" (p. 50). Another, slightly less explicit, formulation can be found in (Dummett 1993, 43–46). My focus on this argument is due to Dummett's own assertion that "neither the objection arising from the manifestation nor that arising from the acquisition of the knowledge is central. The central objection is the circularity of a truth-conditional account" (2006, 55). The discussion of circularity is coupled with discussion of acquisition in this section; manifestation is briefly discussed in footnote 12.

$S \to \exists x \neg\phi(x)$. Dummett grants that if the supertask is executed, then either $\forall x\phi(x)$ or $\exists x\neg\phi(x)$ will become TIC, i.e. he accepts

$$S \to \forall x\phi(x) \vee \exists x\neg\phi(x).$$

But from this the disjunction of the two previous formulae,

$$(S \to \forall x\phi(x)) \vee (S \to \exists x\neg\phi(x)),$$

does not follow – not according to intuitionistic logic. Whether $\forall x\phi(x)$ or $\exists x\neg\phi(x)$ will become TIC might depend on specific circumstances about the execution of the supertask. To assume that it is determined in advance which one, is to assume that there is already a truth of the matter in advance of actually going through with the supertask. And that is exactly what is at issue. The classical logician has mounted a circular defense and assumed that there is some categorical, Platonic truth in which to ground the truth-value of the subjunctive conditional.

Let us turn to evaluating this complex argument for intuitionistic logic being the strongest meaningful logic for mathematics. The overall premise that meaning must be learnable is, I think, uncontroversial.[5] So we have to get our hands dirty on the details of the sub-arguments.

The second sub-argument connects back to issues discussed in Chapters 1 and 2. If it is not determined in advance, by pre-existing facts, what the result of executing the supertask will be, that does not necessarily lead to failure of bivalence. However, for two reasons, that is not very important here. First, even if bivalence can be upheld, the classical logician does not get what she hopes for. The alternative suggested in Chapter 2 was that $\forall x\phi(x)$ should be true if present facts makes it necessary that each instance would become TIC, and false if not. This implies that if it is neither necessary that each instance would become TIC, nor necessary that the negation of one of the instances would become TIC, then $\forall x\phi(x)$ is false. This is again consistent with an execution of the supertask actually resulting in all of the instances becoming TIC while we have counted the sentence as false. That is not what the classical logician was bargaining for. Second, and more importantly, Dummett in this case grants the classical logician more than I think we should, namely that

[5]The idea of division of linguistic labor (Putnam 1975) does not conflict with this premise, since it does not challenge the claim that for every expression, *someone* must have learned its meaning. Kripke's (1980) causal theory of the meaning of proper names does imply that there could be names for which it is impossible to identify the referent, because there are no longer any traces of the "initial baptism". However, that does not imply that the *meaning* cannot be learned. And, at any rate, the semantics of proper names is not relevant to the present discussion.

supertasks are metaphysically possible. Being unwilling to accept that possibility, I have to agree with the *conclusion* of the second sub-argument: relying on a justificationist basis is not a viable route to a logic that is stronger than intuitionism.

The problem is instead with the sub-argument concerned with truth-conditional semantics, more specifically the second premise, that truth-conditions are given by sentences of the form T_P. It is true on one interpretation, but not on the interpretation that would make the argument valid. It is true that if someone were to write down an explicit truth-conditional meaning-theory for English, it would contain or imply the sentence T_P for each English sentence P. But it is not necessary for a speaker of English, in particular a child learning the language, to explicitly know T_P in order to understand P. The truth-conditions are "given" to the theorist in the T_P form but do not have to be so "given" to the language user.

Dummett confounds the distinction between, on the one hand, theoretical and practical knowledge and, on the other, explicit and implicit knowledge.[6] The knowledge needed to understand a truth-conditional language is theoretical in that it is about what has to be impersonally satisfied for P to be true, while a justification-conditional language requires practical knowledge from its speakers, i.e. knowledge about what the speaker has to accomplish in order to be in a position to assert P. However, this *aboutness* is not relevant to the argument. What is relevant is what sentences a language user needs to understand prior to understanding P, and T_P is not among them. It suffices to have the implicit grasp of T_P that consists of being able to correlate the left-hand side and the right-hand side.

The sentence just before this one is prone to misunderstanding: in order to correlate the left-hand side and the right-hand side of T_P does the speaker not still need to understand both "'P' is true" and P before being able to understand P, implying non-well-foundedness? This possibility of misunderstanding is due to the difficulty of formulating the T-schema; or rather the difficulty of formulating that version of the T-schema which is central to truth-conditional meaning-theories. The difficulty has to do with the fact that this version is intended to relate language to a non-linguistic reality, but for purposes of discussing it we have to formulate it all in language.

Instantiating the variable P with "the star has five vertices", and assuming we are in a context where the denotation of "the star" is clear, we can distinguish three things: 1) The sentence "'the star has five vertices' is true", 2) the sentence "the star has five vertices" and 3) the state of affairs that the star has

[6]See in particular (Dummett 1993, 46) and (Dummett 2006, 57–58).

five vertices. To emphasize the non-linguistic character of 3, let us symbolize it with the pictogram "$\bigstar$". What a speaker of a language has to be able to relate in order to understand the sentence "the star has five vertices" is 2 and 3. We could write it like this:

($\mathrm{T}^1_\star$) *The star has five vertices* iff $\bigstar$

The problems with this formulation are that $\mathrm{T}^1_\star$ is ungrammatical in the use of "$\bigstar$" and that the distinction between meta- and object-language has only been hinted at with the use of italics and not made explicit. If we have to be both correct and explicit, we have to formulate the same thing like this:

($\mathrm{T}^2_\star$) "The star has five vertices" is true iff the star has five vertices

And now it may seem like we are relating 1 and 2 instead of the intended 2 and 3.

The only requirement for understanding the sentence "the star has five vertices" according to truth-conditional meaning-theories is that the language user can correlate (I use this imprecise word in order to avoid the phrase "is true" and again risking a confusion of 1/2 with 2/3) that sentence with $\bigstar$, i.e., he must be committed to having the same epistemic and doxastic attitudes towards both sides of the biconditional. He does not have to understand the sentence $\mathrm{T}^2_\star$ including the biconditional and the words "is true". He needs implicit knowledge of $\mathrm{T}^2_\star$, but that implicit knowledge can be gained long before the explicit knowledge that $\mathrm{T}^2_\star$ itself is true. The implicit knowledge is typically gained at the age of 3 or 4, I guess, while the explicit knowledge is only attained when the speaker is able to reflect on his own language use – perhaps, indeed, not before he takes a course in the philosophy of language. There is no circularity here to threaten well-foundedness and learnability.

Of course, rejecting this charge of circularity does not bring us all the way to having established that it is possible to convey $\bigstar$ to a child learning English, which it has to be for a truth-conditional theory to be correct. How it actually happens is a highly non-trivial question for cognitive science, but *that* it happens draws overwhelming prima facia support from the fact that almost everyone would claim to grasp it (Dummett being one of the few exceptions, even though his extensive discussions of truth-functional semantics strongly suggest that he actually does understand it, making his writings about it come dangerously close to constituting a performative contradiction). The burden of proof is certainly on Dummett and he is not very successful in shouldering it. In particular, contra (Dummett 1993, 15) it is *not* presupposed that it is possible to entertain the *thought* that $\bigstar$ obtains prior to learning the *sentence* "the star has five vertices". It would be a controversial assumption, and therefore fair game for Dummett, to presuppose that it is possible to think independently of language. But it is perfectly consistent with truth-conditional semantics and the demand for well-foundedness that the understanding of the sentence and the grasp of ★ is attained simultaneously by the child.

Reading Dummett, one gets the impression that he is implicitly subscribing to a *tabula rasa* paradigm of the human mind, so that a child can only come to grasp what can explicitly be explained to it by a teacher, with no aid of the child's innate mental capacities. Let us assume that a given child has learned to verify that a star placed in front of it has five vertices and *so far* understands the sentence "the star has five vertices" according to this justificationist semantics. We can further assume that the child understands other fragments of English, but also only in a justificationistic way. Then it seems correct to say that it is impossible to explicitly explain a stronger truth-conditional meaning of those sentences to that child. That is, verification-transcendent truth-conditions cannot be *defined* from justification-conditions. But it also seems clear that the child brings something to the table him- or herself.[7] At an early age we form the idea of the permanence of physical objects, i.e., when we see our mother on one occasion and then again later, we come to believe that she existed outside our field of vision in the meantime. That is an act of abstraction from mere sense-impressions that we are somehow able to effectuate, and psychological experiments indicate, according to (Carey 2009, chapter 2), that an assumption of an innate idea of the permanency of objects is necessary to explain that process. Using that power of abstraction, the child should be able to move from the idea of verified five-verticed stars over the idea of unverified five-verticed stars to the idea of unverifiable five-verticed stars, relying only partially on help from a teacher.

I do not doubt that a species of creatures that is only able to understand a language with a justificationist semantics is possible. But believing that that species is the human species seems to be a gross underestimation of the role of abstraction, analogy and innateness in the language learning process.

For these reasons I reject Dummett's argument for intuitionism.[8] That leaves the Brouwerian with a Wittgensteinian defense, to which I will return in Section 4.3.[9]

[7] Chomsky (1971) argues that a child's innate mental capacities must be quite extensive for language acquisition to be possible (in general that is, i.e., quite independently of the special case that is here in dispute).

[8] With that verdict I side with the majority. For like-minded attacks on Dummett, see, e.g., (Williamsom 2007, afterword) and (Devitt 1983).

[9] Dummett also gives a second argument for intuitionistic logic in (Dummett 1991a, chapter 24), concerned with indefinite extendability (see (Heck 1993) and (Clark 1998) for discus-

4.2 Truth as potentiality

Even though the task does not constitute the major challenge he claims, Dummett has left us with a question that needs an answer. Having rejected actual infinity in Chapter 1, we cannot take an understanding of the universal (respectively the existential) quantifier on a par with conjunction (respectively disjunction) and extend that to the infinite case. The infinite is essentially non-extensional, while the child's *first* understanding of a universally quantified sentence surely happens with an example where all the instances are surveyed. How is the understanding of the quantifiers generalized from there?

When a child has learned to understand $\forall x\phi(x)$ in the context of finite domains by checking them all, it is a small accomplishment – the kind of accomplishment I would claim, contra Dummett, that a human child can achieve on its own – to realize that if it is told that some sentence of the form $\forall x\phi(x)$ is true, it can be inferred that if any random object a in the relevant domain is picked out, it must be the case that $\phi(a)$, and conversely that the truth of $\forall x\phi(x)$ follows from it being the case that whenever a random object a is chosen, it must be the case that $\phi(a)$.[10] That is, that $\forall x\phi(x)$ is equivalent to "necessarily, for any object a in the domain, $\phi(a)$".[11] This modal semantics for the universal quantifier is one that is not only learnable, but also gives a stronger logic than intuitionism when generalized to the infinite case, and is not, in the case of mathematics, dependent on Platonism.[12] Consider, in addition to truth-in-content and truth-as-anticipation, a third kind of truth, which we can call "truth-as-potentiality" or "TAP" and give the following definition (to be amended below) together with "false-as-potentiality" or "FAP":

- $P(a_1, \ldots, a_n)$ is TAP if it is possible to make $P(a_1, \ldots, a_n)$ TIC

sion). Indefinite extendability will be discussed in Section 5.9 and on page 207.

[10] As a high level of formal precision is not called for here, I am using "a" both in the object language and the meta-language.

[11] The universal quantifier figures in the explanans, making this explanation circular. This means that the explanation cannot be used to convey understanding of the quantifier to someone who does not already grasp it. But that is not the point; in fact I have already stated that I believe that to be impossible to do through definitions. Instead the project is to supply an ontological underpinning to justify the established use of language, as far as such a justification can be given (it will be clear later that I think some revision is needed when we go beyond arithmetic).

[12] With this semantics, understanding of the universal quantifier can also be manifested, as Dummett demands; see his (1978b, 224–226), (1991b, 314–316) and (1993, 46). Say that person A is convinced (for whatever, perhaps irrational, reason) that the sentence $\forall n\phi(n)$ with a quantifier that ranges over the natural numbers is true. If person B informs A that B has written a natural number N on a slip of paper and asks A if he is willing to bet on N satisfying ϕ with odds $N : 1$ without being told what N is, A manifests his classical understanding of $\forall n\phi(n)$ by accepting the bet. (This example depends on every element of the domain being nameable, but that is an assumption I accept for independent reasons below.)

- $P(a_1, \ldots, a_n)$ is FAP if it is possible to make $\neg P(a_1, \ldots, a_n)$ TIC[13]
- $\phi \vee \psi$ is TAP if ϕ is TAP or ψ is TAP
- $\phi \vee \psi$ is FAP if ϕ is FAP and ψ is FAP
- $\phi \wedge \psi$ is TAP if ϕ is TAP and ψ is TAP
- $\phi \wedge \psi$ is FAP if ϕ is FAP or ψ is FAP
- $\phi \rightarrow \psi$ is TAP if ϕ is FAP or ψ is TAP
- $\phi \rightarrow \psi$ is FAP if ϕ is TAP and ψ is FAP
- $\neg\phi$ is TAP if ϕ is FAP
- $\neg\phi$ is FAP if ϕ is TAP
- $\forall x\phi(x)$ is TAP if necessarily for any possible object a satisfying the criteria for being in the domain, if the sentence $\phi(a)$ is formulated, it is TAP.
- $\forall x\phi(x)$ is FAP if it is possible to formulate a sentence $\phi(a)$, where a satisfies the criteria for being in the domain, which is FAP.
- $\exists x\phi(x)$ is TAP if it is possible to formulate a sentence $\phi(a)$, where a satisfies the criteria for being in the domain, which is TAP.
- $\exists x\phi(x)$ is FAP if necessarily for any possible object a satisfying the criteria for being in the domain, if the sentence $\phi(a)$ is formulated, it is FAP.

There are quite a few aspects of TAP that need to be commented on. I will restrict myself to the case of arithmetic here; the case of set theory is discussed in the next chapter and the rest of the dissertation. Further, the challenge from rule following skepticism will be ignored for the moment and treated in the next section.

Truth-as-potentiality provides for an arithmetic which is equivalent to classical arithmetic. For example, it is either necessary that for each natural number n, if the sentence "if n is an even number larger than 2, then there exist two prime numbers of which n is the sum" is formulated, then it is TAP, or possible to formulate some sentence of that form which is FAP. In the first case Goldbach's Conjecture is true-as-potentiality, in the second it is false-as-potentiality. That this is so relies crucially on the fact that every object in the domain of arithmetic is named, so that every instance of a quantified sentence

[13] Below, the notion of TIC will be revised, and in that connection the dual notion of falsity-in-content (FIC) will be introduced. Then the slight awkwardness of defining FAP of $P(a_1, \ldots, a_n)$ in terms of the TIC of its negation can be avoided: $P(a_1, \ldots, a_n)$ is FAP if it is possible to make $P(a_1, \ldots, a_n)$ FIC.

is itself a sentence. That way, "modal quantification" over sentences that can be formulated is equivalent to quantification over a Platonic domain (as can be seen from a trivial induction).

This equivalence is gained in such a way that no assumption of actual infinity is needed. It is not assumed that all the natural numbers exist, for we just rely on what would be the case for each one if it were to be constructed or its name formulated. It is also not assumed in the case of a quantified sentence that all its infinitely many instances exist, for, again, the truth value is based on what *would* happen if they *were* formulated. And finally, we have not by this recursive definition defined an actual infinity of sentences to be TAP. The clause "$\phi \wedge \psi$ is TAP if ϕ is TAP and ψ is TAP" is short for "$\phi \wedge \psi$ would be TAP if formulated, if ϕ would be TAP, if formulated, and ψ would be TAP, if formulated", and similarly with the other clauses. Existence of sentences is also not to be understood Platonically but simply means that someone has formulated the sentence.

It must also be noted that we gain the equivalence even though we are, just like Brouwer, locating the truth makers in the austere ontology of mental constructions: TAP is, just like TAA, grounded in the possibility of TIC. It is true that we are to a larger extent relying on merely potential mental constructions. But also Brouwer relies to some extent on potential constructions, and we have found no convincing argument to make us believe that the limit of what they can support is given by TAA and not TAP. TAA and TAP are alike in the way that a sentence may be either without being capable of being TIC, because it is about an infinity of cases. The difference between TAA and TAP is that the former requires some foreknowledge on the part of the creating agent that TAP does not. But that extra epistemic requirement is not forced upon us by the choice of a mentalistic ontology. It really is *extra* and, to repeat, we have found no good argument why we should yield to it and not accept the possibility of verification-transcendent truth.

The theory that I have briefly presented here and will develop in the remainder of this dissertation is a constructivistic theory in the strict sense of that word, but not in the sense it has acquired. I am a constructivist in the sense that I believe that the truth makers of a true mathematical sentence must be constructable (individually, not necessarily collectively). According to the established meaning of "constructivism", one must further believe that the constructibility must be known, if a sentence is to be true. Thus the established position should be called "verificationist constructivism", and the more general label should be used in a broader sense.

For reasons that will be discussed in Section 4.4, I claim that mathematics is about temporal processes and that truth-in-content therefore has the strongly temporal character which implies that $2+2 = 4$ was not true in that sense before anyone had constructed it. However, the pre-theoretic concept of truth is more naturally identified with truth-as-potentiality than with truth-in-content, and the former is not temporal to the same degree as the latter. A sentence cannot change its status from being not-TAP to being TAP or *vice versa* in time;[14] if it ever becomes true, it is so from the beginning. TAP is temporal to some degree, though, for on pain of letting in actual infinity through the back door, we must admit that *sentences* come into existence at a point in time, namely when someone formulates them. Thus a sentence which is TAP began being TAP at a point in time: the point in time when the sentence itself came into existence.

Why is it so extremely counterintuitive when Brouwer claims that $2 + 2 = 4$ was not true before it had been proved? The natural reaction to the claim is: "But *if* someone had calculated $2 + 2$ prior to that point in time, *then* they would have got the result 4". This reaction we can honor with the concept of TAP; it is actually exactly what it means to say that $2 + 2 = 4$ is TAP (except that it is also part of the content that the same would happen at any later point in time). We cannot say that $2 + 2 = 4$ was TAP before the sentence was formulated, and that may still strike some (those who are willing to posit abstract propositions as truth-bearers) as strange, but the core of the intuition is salvaged. The core of the intuition is about determinateness: it was given in advance of anyone calculating $2 + 2$ what the result would be. And that we can acknowledge as constructivists. That mathematical truths are truths about temporal processes does not imply that the truth is temporal in any *significant* sense.[15]

To justify the claims made in this section we need to withstand the challenge from the Wittgensteinian rule-following skeptic. In the following section it is discussed how to do that, after which we return to the subject of TAP and discuss it in more detail.

4.3 Wittgenstein's rule following considerations

We owe Dummett an answer to one more challenge before being justified in accepting TAP as a legitimate notion of truth, namely the one about subjunc-

[14] This is due to *monotonicity*; see Section 4.4.

[15] The sense of counterintuitiveness may be alleviated further by noting that a sentence existing at one point in time may be true *about* another point in time. So just as "it is raining" may be false about today but true about yesterday, $2 + 2 = 4$ *is* true-as-potentiality *about* any point in time.

tive conditionals. Might it not be the case that it is *possible* for each even natural number larger than 2 to construct it together with two prime numbers and get the sum of the latter two to be equal to the former, but also *possible* to construct an even natural number larger than 2 in such a way that it cannot be made equal to two primes? If so the Goldbach Conjecture would be FAP even though each instance of it is potentially TIC, making the expression "the Goldbach Conjecture is FAP" strongly misleading: the bivalence that is built into the clauses for TAP and FAP (conditional on bivalence for atomic sentences) would be deceitful and not reflect determinacy in the subject matter. In that case, TAP could not reasonably be considered a legitimate kind of truth. For it to be a legitimate kind of truth it has to be the case that, for each even natural number larger than 2, it is given in advance what would happen if someone were to construct it and search through the possibilities for decomposing it into a sum of two primes. The rule following skeptic claims that it is not.[16]

Let us say that we are considering the number 60 and claim that it does have the Goldbach property, for it is the sum of the two primes 7 and 53. We may then be challenged by a skeptic who claims that it was not necessary that we got 60 as the result of adding 7 to 53, we might just as well have reached the result 61. We would like to answer that if we had reached that result, it would be because we had not applied the addition function, but some other function, or just picked a number at random, and that it is necessary that a calculation of the result of applying *the addition function* to 53 and 7 would have resulted in 60. However, it is obvious that that does not suffice as an answer; we need to back it up with an explanation of what the addition function is and how it can be grasped by a subject. Given that it is an infinitary object and we have forsaken the option of locating it in a Platonic realm, the skeptic may well think that we are unable to provide such an answer. The first part of our answer is that the function is not to be identified with an infinity of triples as the classical mathematician does, but is rather to be understood as a finite rule, i.e. intensionally rather than extensionally.

But that answer just induces the skeptic to bring out his Wittgensteinian arsenal and aim it at showing that as finite beings we cannot have a unique infinite function in mind when we think of "the addition function". We have only ever gone through a finite number of instances, so they in themselves cannot commit us to one specific way of answering all future queries about

[16]The original sources are, of course, (Wittgenstein 1953) and (Wittgenstein 1956), but I am relying heavily on (Kripke 1982) and (Wright 1980). Another important reference is (Miller and Wright 2002). I am ignoring what may be interpreted to be Wittgenstein's own solution to the skeptical problem in §201 of (Wittgenstein 1953).

addition, as there are infinitely many different ways of extrapolating. If we have never before added 7 and 53, what is it that determines that only 60 could be the correct answer as we understood “addition” before being confronted with that specific example? Specifically, why could we not just as well have meant *quaddition*, defined by

$$n \oplus m = \begin{cases} 5 & \text{if } n = 7 \text{ and } m = 53 \\ n + m & \text{otherwise,} \end{cases}$$

so that 5 is the answer.

We might answer that it is commitment to following the rules that define addition, encoded in the equations $n + 0 = n$ and $n + m' = (n + m)'$, where n' denotes the successor to n. But those rules also have infinitely many potential instances and only finitely many constructed instances, so the skeptic can ask virtually the same question again, this time about the meaning of “successor” and of the equality sign. Likewise, it seems that any other answer can prompt an equally good question from the skeptic and thus does no more than start a regress. See the literature mentioned in the last footnote for rehearsal of a large range of such answers. Only those that are closest to my preferred answer below, and therefore interesting to compare with, will be discussed here.

If we have to conclude that rule following skepticism is correct, then there might be a Wittgensteinian argumentative strategy available to get from that conclusion to the additional Brouwerian conclusion that TAA is the weakest (i.e. has the largest extension) legitimate notion of truth.

Let me explain using the example of the law of associativity, $\forall k \forall n \forall m(k + (n + m) = (k + n) + m)$, which is TAA, but was not prior to being proved. It can be proved by applying rules of inference to the definitions of addition (as given above) and of natural numbers (as given by Dedekind (1888) or Peano (1889)). So according to a realist, or, better, “a rule-following believer”, the law of associativity is implicit in those definitions. That is, accepting those definitions carries with it an implicit commitment to accepting the law, a commitment which is brought out by producing the proof but was there all along. The rule-following skeptic denies this. The skepticism towards rules applies just as much to the rules of inference as to the rules of addition. Therefore, the “commitment” to the inference rules of (classical or intuitionistic) logic could have been interpreted in such a way that they did not yield the law of associativity when applied to the definitions of addition and natural numbers. Hence, before inventing the proof we now have, there was no sense in which $\forall k \forall n \forall m(k + (n + m) = (k + n) + m)$ was true.

Since the day the proof was invented and was accepted by the mathematical community, it has been true, in a particular Wittgensteinian way. First, it has been true that $\forall k \forall n \forall m(k + (n + m) = (k + n) + m)$ does follow from the definitions and the inference rules, for by accepting the proof, the meaning of those inference rules has been modified: now the specific instances of the rules that are used in the deduction have become a part of the meaning of the rules. Second, the law of associativity has now become a new standard for correct application of the rules of addition. Before, I could calculate $k + (n + m)$, convinced that I followed the rules of addition and not being answerable to an objective standard, and get one result, and calculate $(k + n) + m$, convinced that I followed the rules of addition and not being answerable to an objective standard, and get another result. Now, the theorem sets the standard that I must reach the same result in the two calculations; if not, the community will judge that I must have committed an error in my attempts to follow the rules.[17]

This outlook on the role and ontological status of proofs could be used in support of Brouwer. One might argue that if neither the truth nor the falsity of a theorem follows "by itself" from definitions and axioms, then it is correct to repudiate bivalence. And one might find it reasonable to insist on constructive proofs of existential claims for the same reason: a truth maker for the claim is not determined by the definitions and axioms themselves, so only if we have a method for finding one is the claim justified.

However, Wittgenstein's rule-following skepticism is so corrosive that it could also be used to undermine Brouwer's position. It is algorithms which are the truth makers for sentences that are TAA, and algorithms are just a special kind of rules. They do not, according to Wittgenstein, objectively determine an outcome of executing them. That is, they do not objectively determine what can potentially be a TIC-maker for the sentence. Accepting a sentence as TAA is just to decide to count an execution of the algorithm that does not result in the production of a TIC-maker as a misexecution.[18]

Then again, in some places[19] Wittgenstein refuses to be revisionary towards

[17]I or the community may of course give this new standard, which is itself a rule, a deviant interpretation in any specific case and declare that the results 5+(2+4)=11 and (5+2)+4=12 do not constitute a violation of the law of associativity. Also the new standard has no more force than what is attributed to it.

[18]Brouwer seems to try to take a third, intermediate position where there are only two. There are only two possible positions regarding what happens when a subject constructs the elements of a lawlike sequence: either he invents the elements (perhaps *feeling* that his hand is tied) or he discovers something predetermined. The predetermination of the decimals of π implies the predetermination of k_1. If Brouwer denies predetermination as I have explicated it (perhaps just referring to the subjective feeling with the word "predetermined" – see footnote 2 on page 63), he cannot allow TAA that goes beyond TIC for that involves rule-following trust (unless he understands the "true" in "true-as-anticipation" in the special Wittgensteinian way, and he certainly does not).

[19]For example (Wittgenstein 1956, V, 52).

classical mathematics because its theorems actually are true in the strongest sense that Wittgenstein would allow a mathematical theorem to be: they are accepted by (the vast majority of) the mathematical community as a standard for judging when rule-following has gone wrong.

I will not go deeper into a discussion of which conclusion a rule-following skeptic should draw.[20] I will instead argue for trusting in rule-following and then proceed by investigation the consequences of this trust. Let me just explicitate what can be concluded from the discussion so far: It is not clear that rule-following skepticism can be used successfully in support of the thesis that TAA is the strongest legitimate notion of truth. It was noted in Chapter 3 that Brouwer requires actual constructions in some cases while being happy with potential constructions in other cases. The problem of justifying this is just moved back one step if rule-following skepticism is brought in in defense: some rules (rules with infinitely many steps, e.g. the one for the construction of π) are distrusted while other rules (finite algorithms) are trusted.

Let us first consider whether the Platonist is in a better position to withstand the skeptical challenge. Is there a plausible answer which is only available if we accept Platonism and not if we keep to constructivism? Kripke (1982, 53–54) answers this question in the negative, and I see no reason to challenge him in that respect. The Platonist claims that the addition function exists as an actually infinite set of (something like) ordered triples in the Platonic realm and that this set is the truth maker of truths about addition. However, the existence of this set does not in itself provide a solution to Wittgenstein's challenge which is to account for how a subject is able to refer to addition. The Platonist has the referent, but she still needs an account of the connection between the referring expression or thought and that referent. If there is a Platonic + then (we must assume in the absence of an argument to the contrary that) there is also a Platonic $\oplus$. How do we manage to refer to the former rather than the latter with the word "addition", the skeptic asks? The answer has to be located outside the Platonic realm, in the mind of the language user or in the language community, just as before.

So the question of rule following skepticism versus rule following trust is independent of the question of Platonism versus mentalism. Rule following is about creating a connection between a finite rule as understood by a mathematician and its infinite extension. This problem does not go away when it is assumed that the infinite extension is a Platonic entity.[21]

[20] The reader is referred to part 2 of (Wright 1980) where the questions is treated in much more detail.

[21] It also does not help to posit, as Frege (1892) did, objective, abstract senses to facilitate the connection between concepts and referents, i.e. supplementing mathematical Platonism

I believe that the correct – and Platonism-independent – answer to the skeptic is, in a word, simplicity. On that issue Kripke (1982, 38, original emphasis) writes the following:

> Let no one – under the influence of too much philosophy of science – suggest that the hypothesis that I meant plus is to be preferred as the *simplest* hypothesis. I will not here argue that simplicity is relative, or that it is hard to define, or that a Martian might find the quus function simpler than the plus function. Such replies may have considerable merit, but the real trouble with the appeal to simplicity is more basic. Such an appeal must be based either on a misunderstanding of the skeptical problem, or of the role of simplicity considerations, or both. Recall that the skeptical problem was not merely epistemic. The skeptic argues that there is no fact as to what is meant, whether plus or quus. Now simplicity considerations can help us decide between competing hypotheses, but they obviously can never tell us what the competing hypotheses are. If we do not understand what two hypotheses *state*, what does it mean to say that one is 'more probable' because it is 'simpler'? If the two competing hypotheses are not genuine hypotheses, not assertions of genuine matters of fact, no 'simplicity' considerations will make them so.

I think that Kripke is completely correct in his critique of the specific way to use simplicity as a solution contemplated here, so this is not the solution that I will advocate. But there is another way to use it, where simplicity comes in earlier, so to speak, not considered by him. In addition to the epistemic role that simplicity can play in the choice between hypotheses, I claim that simplicity can play a constitutive role for a rule. A subject can succeed in meaning $n + m' = (n + m)'$ with "$n + m' = (n + m)'$" if he commits to a finite number of examples such as $0'' + 0'' = (0'' + 0')'$, $0''' + 0'''' = (0''' + 0''')'$ and $0+0''' = (0+0'')'$ and also commits to extrapolate from these examples in future cases *in the simplest possible way.* I further claim that we have an inborn sense of simplicity which is partially constitutive of rationality which we are able to apply when needed, independently of an ability to articulate general rules for when something is simpler than something else. (So this claim is a natural continuation of the reply to Dummett above.)

Let us consider this proposal in relation to Kripke's objections, starting with the "real trouble": Kripke only argues that *if* we have not succeeded in referring

with linguistic Platonism, for that just pushes back the problem a step: how do our physical words and mental thoughts connect with (one specific entity in) *this* abstract realm.

to a specific hypothesis, then we cannot use simplicity epistemologically to give warrant for believing it. It is correct that we need to be able to state a hypothesis prior to figuring out its truth value. But that we have to be able to "pick out" rules/hypotheses *semantically* independently of *epistemically* "picking out" the true hypothesis from a range of options, does not in the least imply that we cannot also use the concept of simplicity in our efforts to do the former.

It is also irrelevant that a Martian might find the quus function simpler than the plus function. All the human subject needs in order to have the ability to mean addition with some expression is that he is in possession of *some* simplicity measure according to which addition is the simplest extrapolation of some finite range of examples. Maybe the Martian's sense of simplicity is such that the Martian needs a different, larger range of examples to be able to pick out addition. Or maybe the Martian's sense of simplicity is such that no finite range of examples will suffice to do that, in which case only the human and not the Martian will be able to do arithmetic. For humans to do arithmetic and talk about it, it is not a necessary condition that all language-using species can do the same.

This of course generalizes to the issue of relativity of simplicity between humans. In order for all (or almost all) humans to be able to communicate it is sufficient that there are enough "basic concepts" (perhaps including the rule $n + m' = (n + m)'$) from which other derived concepts (perhaps including the rule for addition) can be defined, which are such that for *each* person there is *some* range of examples that will suffice for that person to grasp, with the aid of the person's individual sense of simplicity, the concept. This is compatible with a significant amount of relativity. It is not compatible with those inborn senses of simplicity being completely unrelated, so if the skeptic claims that we all mean something different with the rule "$n+m' = (n+m)'$" and that that would be revealed if we asked different people to apply it to larger numbers than had been done at any previous point in history, we cannot prove him wrong. But fortunately this skeptic is not Cartesian; the rules of the game are different. We are not searching for a way to *know* the correct answer in a dialectic situation where the criteria for knowledge can be strengthened without end. We are asking a semantic question in a dialectic situation where apodictic knowledge is not the goal. This skeptic must accept a plausible answer. He is not free to throw more and more outrageous skeptical scenarios at us just because he can think of them. I do believe that those inborn senses of simplicity have a sufficient overlap and that that overlap is a significant part of the explanation for our ability to communicate.

Finally, *defining* simplicity is not needed in order to make the appeal to it that is made here. It is claimed to be a primitive human faculty, so the language user needs no explicit knowledge about it. If explicit knowledge can be gained by a philosopher of science – and Kripke is probably right when he writes that that would be difficult to achieve – it would be impossible to convey it to someone who did not already have implicit knowledge about simplicity. For the definition of simplicity would need to be formulated in terms of rules, and a person lacking that implicit knowledge would not be able to interpret them. This is analogous to the case with the T-schema above: an implicit knowledge of both simplicity and the idea of truth conditions has to come before explicit knowledge of the same.[22]

Connecting back to Chapter 2, my claim can be formulated as follows: at a given time, t_1, the rule $n + m' = (n + m)'$ may for a given subject be identified with

$$\langle 0''{+}0''{=}(0''{+}0')', 0'''{+}0''''{=}(0'''{+}0''')', 0{+}0'''{=}(0{+}0'')', \\ \text{intention to expand only in the simplest possible way}\rangle,$$

and may later, at t_2, be identified with

$$\langle 0''{+}0''{=}(0''{+}0')', 0'''{+}0''''{=}(0'''{+}0''')', 0{+}0'''{=}(0{+}0'')', 0{+}0''''{=}(0{+}0''')', \\ \text{intention to expand only in the simplest possible way}\rangle,$$

and the numerical identity between them does not rely on a decision at t_2, but obtains in virtue of the subject's sense of simplicity as it exists already at t_1.

The cognitive process described here will of course not normally take place at the conscious level. Explicitly, a human mathematician will simply commit to "following the rules" or even more simply "do mathematics". It is in principle possible for this cognitive process to be conscious, but I doubt that it ever actually is: anyone with even the most basic mathematical training will be so used to dealing with syntactic rules that they will have the feeling that they understand the rules as written in mathematical textbooks without any mediating means, and a child just beginning to learn mathematics does not reflect on what commitments they accept. But it is principally impossible that the simplicity measure is the result of a choice on the part of the subject, for any choice would have to be interpreted in future applications and for that a more basic simplicity measure would be needed; it must be instinctive.

[22] There is a disanalogy: In the present case it is strictly speaking wrong to talk about implicit *knowledge*, since, as argued above, there is no need for an objective simplicity measure to have knowledge about. It would be more correct to say that the language user needs to have an implicit *criterion*.

It is instructive to compare this simplicity answer with another answer discussed by Kripke, namely the dispositional, according to which meaning addition by "addition" is to be disposed, for any natural numbers n and m, to answer with the sum of n and m when asked to add n and m. Kripke criticizes this in two ways. First, he points out that no human being is disposed to answer with the sum of n and m in more than a finite number of cases, for if the numbers are sufficiently large, we are unable to do the computation. Second, with this answer we cannot account for the concept of making a mistake in a calculation – whatever result we in fact reach must be what we meant to be the sum – and that does not square with our intention to submit to a norm when we make a claim about what the sum of two numbers is.

The simplicity answer is not subject to those problems. Where dispositions only cover a finite number of potential cases, my sense of simplicity outruns my practical ability to apply it and "reaches out to infinity". I have a clear sense of what it is to be the simplest answer to a question of the form "what is the result of rewriting $n + m'$ similarly to $0'' + 0'' = (0'' + 0')'$, $0''' + 0'''' = (0''' + 0''')'$ and $0 + 0''' = (0 + 0'')'$?", even in cases where n or m is too large for me to actually answer. That is, I have a clear sense of how I would go about computing the answer if I had more computational power, enlarged memory capacity, etc. And I have imposed a norm on myself that makes it possible to distinguish conceptually between a correct and an incorrect calculation.

One way to formulate the rule following problem is like this: the understanding of a rule has to be immanent to the subject, but the subject is finite and the rule may be infinitary; if the rule is immanent, how can it transcend us? To this formulation my answer is that it can because what is immanent to the subject is the idea about what it is for someone to follow the rule, i.e. an understanding of what is common to all possible applications of the rule, while it is irrelevant what the subject itself actually does (as opposed to "tries to do" or "would try to do") when applying the rule. (Compare: A lame person can have a clear understanding of what it is to run.)

Of course, this is exactly what Wittgenstein is famous for denying, so I by no means claim to have refuted him. I only claim to have an answer that is plausible and consistent with (non-verificationist) constructivism.[23] As, furthermore, the arguments above show that the Wittgensteinian justification of

[23] Wittgenstein's rule following skeptic seems in places, e.g. (Wittgenstein 1956, part 1, paragraph 113), to be even more extreme than Kripkenstein's. He is uncooperative, intentionally trying to find loopholes in the instructions he has been given, so he can claim to follow them but deviate from the intended interpretation, *even when he very well understands what it is.* The behavior of such a troublemaker is quite irrelevant when we are discussing the possibility of having a unique rule in mind. The possibility of intentional misunderstanding does not rule out the possibility of understanding. You can claim to have a completely different sense of simplicity than someone else, but you probably do not, and even if you do, that does

Brouwer, if there is one, is completely independent of the choice between believing that mathematical objects are mental constructions and that they exist atemporally in a Platonic realm, one who has adopted the former position has the same right as the Platonist to develop a mathematics that is based on trust in rule-following.

4.4 Arithmetic

So that is exactly what we will do – and have already briefly started doing in Section 4.2. First, however, it seems prudent to pause to take stock.

It was concluded in Chapter 1 that there is little reason for believing in a Platonic ontology for mathematics and in actual infinity. We therefore turned to Platonism's historically most important alternative, intuitionism, for answers. Hoping to locate a legitimate notion of infinity, we were disappointed in Chapter 2 to find out that lawless choice sequences are not more than forever finite sequences that their creators have an intention to expand. Wittgenstein would pass the same judgment on lawlike sequences, but it has been argued that we can plausibly disagree with him, without smuggling in actual infinity through the back door, by trusting in human beings' ability to grasp a rule. So, having rejected actual infinity and non-rule-based potential infinity, we must conclude that the only available ontological basis for an infinite mathematics is to be found in rules.

These must be rules for what can be realized in this world (which contains physical entities and mental entities – one kind of which may or may not be reducible to the other – but which is not, as far as I am willing to make assumptions, inhabited by extra abstract entities). Thus we will proceed on the assumption that Brouwer, as interpreted in Chapter 3, was partially right: our most liberal concept of truth must be grounded in the concept of truth-in-content.

On the other hand, we did not find convincing reasons in Sections 4.1 and 4.3 to follow Brouwer in thinking that this most liberal concept of truth should be restricted by what is verified or verifiable. Thus the concept of truth-as-potentiality seems perfectly legitimate. In particular, we are able to vindicate classical arithmetic on an austere ontological basis – this was briefly discussed in Section 4.2 and will be treated in more detail in this section.

---

not prevent that someone else from fixing on a unique rule. Another, also more extreme, interpretation can be found in (Wright 1980, chapter II, section 3), but it is dependent on Dummett's conclusion that we cannot understand verification-transcendent truth criteria, which we have already rejected.

Let us try to be more specific about what exactly that ontological basis is. Rules are of central importance. On pain of actual infinity we cannot identify a rule with the extensional set of all its potential applications. And Wittgenstein's rule-following challenge could not be met if we turned to formalism and conceived of rules as merely syntactic objects. Hence, it is rules as intensional objects, and more generally interpreted language, i.e. language as understood by a subject, that we are giving central place in our account of mathematics. This is in opposition to both the intuitionist and the Platonist.

On the one hand we have the rules themselves, and on the other we have the results, actual and possible, of applying those rules. Where, ontologically, should we locate the latter? The rules can be implemented in a Turing machine which can then produce the results. So is mathematics about Turing machines? Appealing to a notion of abstract Turing machines is of course out of the question, so the proposal would have to be that it is concrete, physical Turing machines that is the subject matter of mathematics. Therefore, there is a problem with that proposal, namely that any output from such a machine, e.g. four strokes on a piece of paper as a result of calculating $2 + 2$, contains irrelevant information, such as the length, thickness and color of the strokes, that has to be abstracted away in order to get the result of the rule itself. And that is an act of interpretation, that is, an act that only a mind can effect. Thus it is really the interpretation of the result (and the interpretation of all the steps toward that result) that mathematics is concerned with. We must therefore agree with Brouwer on one more point, namely that mathematics is about mental constructions.[24] A sentence is made TIC by a subject constructing its truth-maker at a given point in time. And – as will become important in the following – the relation of dependency that obtains between certain sentences' becoming TIC is a temporal relation. A conjunction, for instance, can only become TIC *after* both of its conjuncts have.

This does not mean that we have to agree with Brouwer on the specific, rather mystical, details in his account of the mathematical mind. Specifically, there seems to be no compelling reason to accept the requirement that all mathematics must be reducible to the "empty two-ity". Instead, the conclusion we have reached is that all rules that can be grasped with a finite number of ex-

[24]The mental state of the subject making mathematical constructions can contain irrelevant information, just like a physical output, but the point is that the subject's consciousness can be intentionally directed towards just the relevant information and in that way abstract away from contingent aspects of the mental state, if for example the subject tends to do arithmetical calculations by imagining strokes with length, thickness and color. This is a further departure from Brouwer. If I interpret his (1948) talk of the "empty two-ity" and the "deepest home of consciousness" correctly, he is assuming that effecting mathematical constructions that are pure in themselves, so to speak, is within the powers of the subject.

amples and our sense of simplicity are legitimate as basic building blocks.[25] Therefore, we will redefine the concept of TIC below. The adjustment is quite small, though, and most of what has been said about it in Chapter 3 will still hold good.

Rules are also – to finally make that explicit – what we point to in order to answer Dummett's question about subjunctive conditionals. Subjunctive conditionals are at the heart of the notion of truth-as-potentiality: a sentence being TAP depends on what would happen if the subject executed certain constructions. And we must grant Dummett that such conditionals cannot be barely true. We thus do owe him a straight answer, and that answer is that rules are the truth-makers for mathematical subjunctive conditionals. Goldbach's Conjecture is TAP or FAP by virtue of the rules that define "natural number", "even", "sum", "prime" and the logical connectives and quantifiers. (Actually, Dummett demands a categorical *statement* as truth-maker for a subjunctive conditional, and that we must deny him. The content of a basic rule cannot be fully and informatively expressed in a statement as it relies on an essentially implicit sense of simplicity.)

Next, let us consider if there is a finitist challenge that may give problems for our theory of mathematics. It has been suggested[26] that intuitionism is unstable in the sense that the reasons given for intuitionism are more plausibly reasons for strict finitism. A core claim of intuitionism is that a necessary condition for a mathematical sentence being true is that it is verifiable in principle. But the idealization to what is verifiable *in principle* rather than just verifiable given the actual constraints imposed by the shortness of human life, limited memory, etc. may seem like an unjustified element of realism. The thought is that the idealizing intuitionist is appealing to what a subject which is just like a human, just with a longer life, more memory, etc. would be able to verify, and to assume that what such a being is able to verify is already determined and independent of the specifics of exactly *how* this being is equipped with super-human powers is to assume that there are mind-independent facts of the matter. Therefore, the argument goes, it is only

[25] The traditional desire to reduce mathematics to rules that are as simple as possible can, however, be justified on the present proposal. The simpler the rules are, the higher is the likelihood, presumably, of different subjects succeeding in actually grasping the *same* rule when they try to. Also, an individual may be under the illusion that he grasps a complicated rule only to find out, when needing to apply it in an unforeseen case, that he is unable to decide between two or more ways of doing it. (Unable, not merely in the sense that that he has insufficient memory, time, etc., but in the sense that he realizes that he had not implicitly decided between the options in advance.) Thus, reduction to simple rules serves to minimize the risk of rule following failure.

[26] See the paper "Wang's Paradox" in (Dummett 1978b), chapter VII, section 3 of (Wright 1980) and (Wright 1982). See also (Parsons 1997).

decidable questions *under a certain threshold of complexity* for which bivalence holds prior to execution of the decision procedure.

Without needing to evaluate the force of this challenge to the intuitionist, we can consider whether there might be a similar problem for the present proposal. Are we relying on an idealization of human powers? We certainly do need to appeal to a counterfactual situation where we have a potentially infinite life, a potentially infinite memory, etc. so we could effect any of the possible constructions quantified over in the definition of TAP. However, the role allocated to the *actual* human mathematician is a simpler one on the present theory than it is in the case of intuitionism. According to intuitionism, the mathematician must himself deliver the *truth-makers*, for it is considered illegitimate to assume the existence of something external to him that can serve that role. Thus, the determinateness of constructions by a counterfactual super-human extension of himself may be considered within the scope of what it is illegitimate to assume. On the present proposal, on the other hand, the *actual* mathematician only has to deliver *semantic determinateness*, not truth-makers. He has to understand each rule R in such a way that the extension of "correct application of R" is implicitly determined. Whether or not he himself is able to determine if a given application of R is correct, is immaterial; the task of delivering truth-makers is one he can unproblematically delegate to counterfactual versions of himself or to someone else.

There is certainly a significant *assumption* here, namely that the actual mathematician can have a clear sense of what it is for

$$0^{\overbrace{\prime\prime\prime\cdots\prime}^{N}} + 0^{\overbrace{\prime\prime\prime\cdots\prime}^{M}} = (0^{\overbrace{\prime\prime\prime\cdots\prime}^{N}} + 0^{\overbrace{\prime\prime\prime\cdots\prime}^{M-1}})'$$

to be a correct application of the rule $n + m' = (n + m)'$ even when N and M are astronomical and the instance therefore unsurveyable to him. But there is not *idealization* in the form of unclear "in principle"-modifiers and therefore no *instability*; we rely on 1) what a possible subject could construct and 2) what the actual subject actually does in fixing rules semantically.

That is why we can recover classical arithmetic in the way described above. However, this positive result does not generalize very far. One reason is obvious: with only potential infinity we can at most validate theories of classical mathematics for which the domain is countable. In particular, classical set theory has no place in our reconstruction. An alternative theory of collections will be the focus of the rest of this dissertation, beginning from the next chapter. Leading up to that, the rest of this chapter will be concerned with making it clearer exactly *why* our theory of arithmetic is equivalent to classical arithmetic. There are two reasons. The first, which has already been mentioned in passing, is that every element of the domain of classical arithmetic, i.e. every natural number, has a name. In our ontology we have only found space for completed constructions and rules for finite or potentially infinite, but deterministic, constructions, and both are capable of being individually named. The second is that arithmetic is well-founded, in a certain sense. Explaining what this sense is presupposes a little stage setting, namely explicitly formulating a rule-based arithmetic.

Let for present purposes the language of arithmetic be specified as follows. Terms are what can be composed in the usual way out of 0 (zero), $'$ (successor), + (addition), $\cdot$ (multiplication), variables and brackets. (The standard conventions for suppression of brackets will be employed.) An atomic formula is something of the form $a = b$ where both a and b are terms. On top of that, define "formula" and "sentence" in the usual first-order way, letting negation, disjunction and the existential quantifier be primitive, and conjunction, the conditional and the universal quantifier be defined. For any formula ϕ, let $\phi(x/a)$ be the result of replacing all free occurrences of the variable x in ϕ with the term a. The notation $\phi(a/b)$ will be used ambiguously to denote any formula that is the result of replacing one or more occurrences of the term a with the term b in ϕ.

Arithmetic is then constituted by the following rules that a subject may impose on himself (a, b and c are meta-variables for terms):[27]

- I may at any time introduce the true sentence $0 = 0$.
- I may at any time introduce a false sentence of the form $a' = 0$ or $0 = a'$.
- If I already have a true (false) sentence of the form $a = b$, I may introduce a true (false) sentence of the form $a' = b'$.
- If I already have a true (false) sentence of the form $a = b$, I may introduce a true (false) sentence of the form $a + 0 = b$ (or $a = b + 0$; henceforth I refrain from mentioning such trivial variants).
- If I already have a true (false) sentence of the form $(a + b)' = c$, I may introduce a true (false) sentence of the form $a + b' = c$.
- If I already have a true (false) sentence of the form $0 = b$, I may introduce a true (false) sentence of the form $a \cdot 0 = b$.
- If I already have a true (false) sentence of the form $a + a \cdot b = c$, I may introduce a true (false) sentence of the form $a \cdot b' = c$.

[27] Alternative rules resulting in the same true and false sentences are possible. These are not claimed to be privileged in any way.

$$0 = 0 : \top \qquad\qquad a' = 0 : \bot$$

$$\begin{array}{c} a' = b' : \top(\bot) \\ | \\ a = b : \top(\bot) \end{array} \qquad \begin{array}{c} a + 0 = b : \top(\bot) \\ | \\ a = b : \top(\bot) \end{array} \qquad \begin{array}{c} a + b' = c : \top(\bot) \\ | \\ (a + b)' = c : \top(\bot) \end{array}$$

$$\begin{array}{c} a \cdot 0 = b : \top(\bot) \\ | \\ 0 = b : \top(\bot) \end{array} \qquad \begin{array}{c} a \cdot b' = c : \top(\bot) \\ | \\ a + a \cdot b = c : \top(\bot) \end{array}$$

$$\begin{array}{ccc} & \phi : \top(\bot) & \\ \swarrow & & \searrow \\ \phi(a/b) : \top(\bot) & & a = b : \top \end{array}$$

Figure 4.1: Arithmetical construction rules

- If I already have a true (false) sentence of the form $\phi(a/b)$ and a true sentence of the form $a = b$, I may introduce a true (false) sentence of the form ϕ.
- I may not introduce true or false atomic sentences of arithmetic beyond what is sanctioned by the above rules.

These rules can be represented graphically as in Figure 4.1, using the symbol $\top$ for truth and the symbol $\bot$ for falsity. The notation is to be read as follows: if and when the truth(s)/falsity below are available (both of them in the case of the bottom-most rule), the truth/falsity above can be created. An example, showing how $2{\cdot}3 = 6$ can be made TIC, is displayed in Figure 4.2. Each sentence in this tree can be made TIC by a subject whenever he has already made the sentences below it TIC, and in that sense each truth in the tree depends on the truths below it. It should be clear that the feature of this example, that no sentence depends circularly on itself or on an infinite downward-going sequence of sentences, generalizes to all arithmetical sentences. This is the sense in which arithmetic is well-founded. I will get to the connection between well-foundedness and bivalence in a moment.

To the rules already introduced we add these logical rules, which are represented graphically in Figure 4.3:

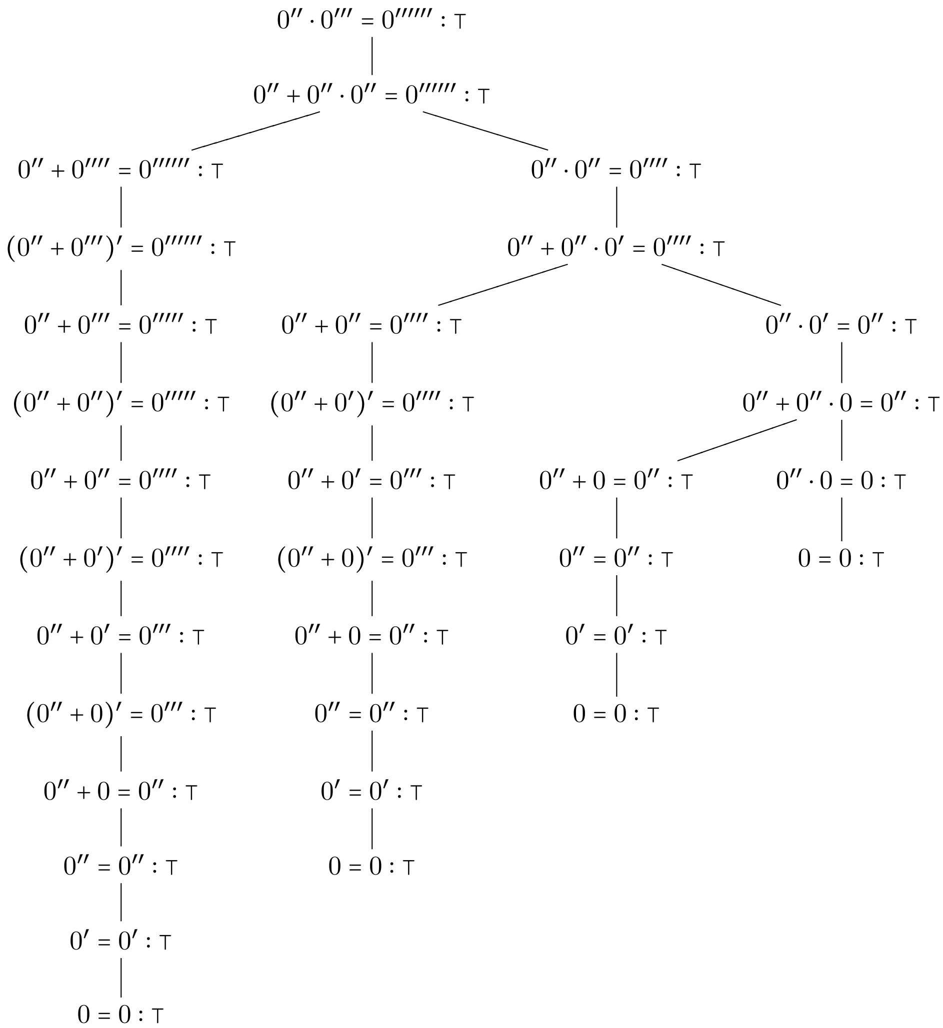


Figure 4.2: The construction of the TIC of $2 \cdot 3 = 6$

- If I already have a true (false) sentence ϕ, I may introduce the false (true) sentence $\neg\phi$.
- If I already have a true sentence ϕ (or ψ), I may introduce a true sentence $\phi \vee \psi$.
- If I already have false sentences ϕ and ψ, I may introduce a false sentence $\phi \vee \psi$.
- If I already have a true sentence of the form $\phi(x/a)$, for some term a of the form $0'''^{\cdots}{}'$, I may introduce a true sentence of the form $\exists x\phi(x)$.
- If I already have a false sentence of the form $\phi(x/a)$ for every term a of the form $0'''^{\cdots}{}'$, I may introduce a false sentence of the form $\exists x\phi(x)$.
- I may not introduce true or false complex sentences of arithmetic beyond what is sanctioned by the above rules.

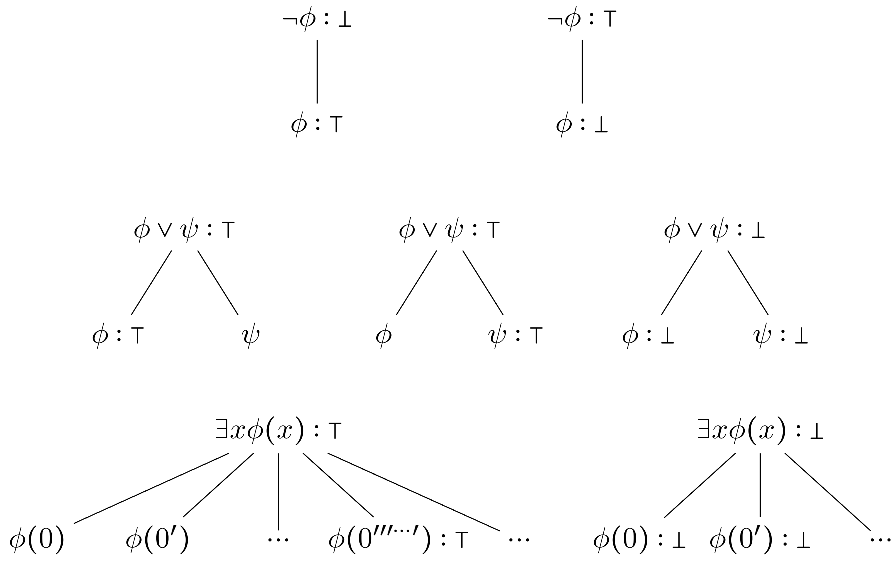


Figure 4.3: Logical construction rules

As a complex sentence only depends on sentences of lower complexity (suitably defined), well-foundedness is intact after the addition of these rules.

What the rules do is that they fix the meaning of the vocabulary of arithmetic by stipulating truth-conditions for the sentences in which it appears. More precisely, they fix the TIC-conditions: when the subject "introduces a true sentence" according to the rules, he makes it TIC. The rules *directly* fix the TIC-conditions. As TAP is defined in terms of what is possibly TIC, they also indirectly fix TAP-conditions.

Notice that in the leftmost rule for disjunction in Figure 4.3 there is no truth value for ψ. This means that the truth value does not matter. ψ can be true-in-content, false-in-content or neither, and in all cases $\phi \vee \psi$ is true. Similarly, in the rule for existential quantification being true: as long as one instance is true, the rest can be true-in-content, false-in-content or neither – in the last case perhaps the instance does not even exist as a sentence (which at any given time must be the case for all but finitely many instances).

It is clear that bivalence fails for TIC, that is, there is a third option in addition to a sentence being true-in-content and false-in-content (FIC).[28] It is not immediately clear that this is a failure of bivalence in a deep sense (just as it was argued in Chapter 2 that the failures of bivalence due to lawless choice sequences were not as deep as Brouwer made them out to be), for there is no

[28]The introduction of the dual notion of FIC does not signal any further departure from Brouwer. It is merely a matter of simplicity of presentation. The discussion in Chapter 3 was most elegantly made in terms of TIC alone, but in the sequel the dual notion will simplify matters.

problem in applying an exclusion negation here and saying that, for any mathematical sentence, either it is TIC or it is not TIC. But the last word on the issue of bivalence has not been said yet; we will discuss it in Chapter 7. What should be noted, though, is that whether or not this failure of bivalence is deep, it is legitimate to have a predicate for sentences for which a construction has been effected that rules out the possibility that the sentence can become TIC, and that is exactly what FIC is.

We want the concepts of TIC and FIC to be monotonic in the sense that for any "possible worlds" W_1 and W_2, if every mathematical construction that has been effected in W_1 has also been effected in W_2, then every sentence that is TIC (FIC) in W_1 is also TIC (FIC) in W_2. Again, there are clearly legitimate and more inclusive notions of truth and falsity according to which "ξ is true" (where ξ is some mathematical sentence) changes from being false to true, so that these notions are non-monotonic, but such contingent matters are of little *mathematical* interest. In order to get "as close as possible" to classical mathematics, we are interested in defining notions of TAP and FAP that are atemporal, if there are any such legitimate notions, and as we have seen, that is exactly what we get (with only the slight caveat that TAP and FAP are as temporal as the existence of the sentences to which they apply) when they are defined as above, relative to monotonic concepts of TIC and FIC. Therefore, we should follow Brouwer in employing such monotonic notions. Not because we are forced to, but because it is the suitable way to abstract from empirical matters that are irrelevant to mathematics.

Given this desire for monotonicity, the logical rules described above present themselves as very natural ways to deal with the *gap* between TIC and FIC. First, they do indeed satisfy the requirement for monotonicity: making $\phi \vee \psi$ TIC on the basis of ϕ being TIC alone can be done without the risk that its truth value may change when more constructions are carried out. Second, they are the strongest possible such rules, and there seems to be no reason why we should go with something weaker and, for example, refuse to make $\phi \vee \psi$ TIC when ψ is neither TIC nor FIC. Thus, we arrive at the Strong Kleene[29] rules for the connectives and the quantifier, and the connection between constructivism and Kripke's theory of truth should now begin to become clear.

However, it is not fully correct to say that Strong Kleene is the strongest possible set of rules that are consistent with monotonicity. Rather, Strong Kleene is the strongest possible *compositional* set of rules that are consistent with monotonicity. The search for suitable rules that are stronger than Strong

[29] See (Kleene 1952, 334).

Kleene will be a major theme later in this dissertation. For the moment we will stick to them.

The modification of the concept of TIC has not changed the fact that there are sentences ϕ such that neither ϕ nor $\neg\phi$ can become TIC (equivalently: there are sentences ϕ such that ϕ can neither become TIC nor FIC). The simplest example is that an existentially quantified sentence cannot be FIC, when, as in the case of arithmetic, the domain is infinite. This of course makes the definition of FIC of an existentially quantified sentence moot, but I have included it anyway in order for TIC and TAP to correspond to each other, in the sense that if I am wrong and completed infinite constructions are metaphysically possible, then a sentence is possibly TIC if and only if it is TAP, and possibly FIC if and only if it is FAP.

For every sentence of arithmetic there is a tree for it, like the one in Figure 4.2, that has the sentence in question at the top and bottoms out in sentences of the form $0 = a$ that can be made TIC or FIC at any time, i.e. sentences that do not depend on further sentences. Such a tree depicts the dependency structure for the sentences in it. As already noted, the meaning of "dependency" can, in the case of TIC and FIC, be cashed out in temporal terms: a sentence ϕ depends on the sentences below it in the sense that ϕ can only be made TIC or FIC *after* the sentences below it have been given in-content truth values (or some of the sentence below it, in the case of TIC of disjunctions and existentially quantified sentences). TIC and FIC travel up trough the tree in time, so to speak.

In a derived and more metaphorical sense, TAP and FAP also travel up through the tree. Although a sentence high up in the tree can become TAP or FAP prior to all the sentences below it getting as-potentiality truth values (because that just requires that it is formulated prior to them), the upper sentence does depend on the (potential) truth values of the lower.

Because each atomic sentence of arithmetic has a finite tree that bottoms out in sentences that can be made either TIC or FIC at any time, it is possible for each atomic sentence to become either TIC or FIC. So as *every* sentence of arithmetic has a tree where every branch passes through an atomic sentence, every sentence that has been formulated is either TAP or FAP. That is how the well-foundedness of arithmetic is responsible for ensuring that the failure of bivalence for TIC does not spill over into failure of bivalence for TAP.

Let me finish this chapter by addressing a potential worry. It may seem like the present account of arithmetic is relying very heavily on objectivism (or, in the sense in which Dummett uses the word, realism) about alethic facts, for instance when the truth value of Goldbach's Conjecture is both claimed to be settled at present even though unknown and made to rely on what it is possible to construct. (Goldbach's Conjecture is TAP if necessarily, for any possible even number n larger than 2, it is possible to construct prime numbers n_1 and n_2 such that it is possible to make $n = n_1 + n_2$ TIC; and it is FAP if it is possible to construct an even number n larger than 2, such that necessarily, for any two prime numbers n_1 and n_2 that are constructed, it is possible to make $n = n_1 + n_2$ FIC.) One may object that this reliance on objective modal facts to partition the sentences of arithmetic into true and false ones is not that different from Platonism.

This is to get the "division of labor" between objective modal facts and the rule-formulating subject wrong. Consider the (potentially infinite) collection of all finite trees of sentences of arithmetic. The alethic assumption that I am making is merely that every such tree is possible, and that is just the relatively minor assumption that there is no finite upper bound on the size of trees that are metaphysically possible (that is, the denial of strict finitism). From there, the subject takes over: it is his ability to semantically pick out a rule that does the real work, i.e. partitions this collection of trees into those that are in accordance with the rules and those that are not. Thus, the formulation "it is impossible to construct a true sentence of the form $k+(n+m) \neq (k+n)+m$ in accordance with the rules" may give the wrong impression, while the equivalent assertion "all possible constructions of a true sentence of the form $k + (n + m) \neq (k + n) + m$ are in conflict with the rules" has the right emphasis. It is of course (to repeat) also a significant assumption that the subject succeeds in partitioning the potential infinity of trees, but as it makes mathematics immanent instead of transcendent, it is not Platonic in nature.

Chapter 5

Classes and real numbers

5.1 Sets, classes and species

We will now turn from arithmetic to set theory – or rather, for reasons to be explained, "class theory" – and stay with that subject for the remainder of the dissertation. We first consider the classical concept of sets, the classical concept of classes and Brouwer's concept of species. Finding them all lacking, we then turn to an alternative.

With the constructivist ontology we can make good sense of the finite levels of the cumulative hierarchy of sets, discussed in Chapter 1. We could identify a set with a thought that "picks out" some sets that have already been so thought. The empty set would then be a thought about nothing. When that has been created, the singleton of the empty set can be created by thinking of that thought. Later, the two-element set consisting of the empty set and the singleton of the empty set can be created by thinking of the two previous thoughts. And so on. No matter how many sets have been created, the subject can always go though them in thought, individually selecting or deselecting each one to create a new set, thus validating the combinatorial conception of collection characteristic of the classical notion of set. Also here, the idea of "dependency" can be cashed out temporally.

However, it is *only* the finite levels of the hierarchy that can be made sense of combinatorially. It is not possible to create an infinite set. Thus this approach would leave us with a highly impotent theory of collections.

Staying classical, there is an alternative to the combinatorial conception, namely the logical conception, where a collection, now called a "class", is something characterized by a criterion of membership: for each open sentence ϕ (of a suitable language) with a single free variable there is a class of the objects that satisfies ϕ. As already discussed in Chapter 1, incautious use of classes leads to paradox. The established way of being cautious is to forbid classes from having classes as elements altogether. They can only have sets as members and thus, figuratively speaking, form a single layer of collections on top of the cumulative hierarchy.[1]

Placing one layer of classes on top of the finite levels of the cumulative hierarchy would not get us very far. We would then have infinite collections, but only infinite collections of finite collections. And at any rate, the method for avoiding paradox seems *ad hoc*. We would need some story about the ontology of classes that explains exactly why it is they can have one kind of collections as members but not another.[2]

Brouwer's species are, like classes, characterized by a criterion of membership, but are allowed to have other species as elements. He defines species as "properties supposable for mathematical entities previously acquired".[3] This needs a little unpacking.

Properties are, as explained in Chapter 3, constructions that can be "embedded" into some objects and turn out to be impossible to "embed" into other objects. So a species cannot be identified with its extension. First, because it is an object in its own right that has to be constructed at some point in time and does not exist merely in virtue of its extension existing. Second, because Brouwer does not, of course, believe that it is objectively given what objects the property is embeddable into; the limits of truth-as-anticipation applies here as well.

Thus we have half the explanation for the choice of the expression "supposable for". A property is created prior to and can exist independently of truths about what it can and cannot be embedded in. So initially it can just be "supposed" to hold for some objects. Actual facts about what it holds for come later. The second half of the explanation is that Brouwer thinks that a property comes with a domain of objects for which it makes sense to suppose that the property holds, roughly in the sense that the property of being red is supposable for a ball but the property of being happy isn't.

The last part of the definition, "mathematical entities previously acquired", explains what the domain is: a property can only (potentially) apply to objects

---

[1]This is the case in von Neumann-Bernays-Gödel set theory and in Morse-Kelley set theory, see (Mendelson 1997, 225–287).

[2]For a more extensive overview of different notions of collections see sections 1 and 2 of (Maddy 1983).

[3]The quote is from (Brouwer 1948, 1237). See also Brouwer's (1918, 150-151), (1947) and (1952, 142).

that were created before the property.[4] There are thus three phases in the construction of a species, that can loosely be characterized as follows:[5]

1. The objects of a domain are created.
2. The species itself is created.
3. Relations of *holds of* or *cannot hold of* are created between the species and (some of) the objects of the domain.

If a species S_1 is itself to be an element of a species S_2, then S_2 would have to be introduced after S_1. Like the classical mathematician, Brouwer therefore ends up with a hierarchical notion of collections. There are species of order one, i.e. species that only have non-species objects as elements, species of order two that can also have species of order one as elements, species of order three, etc.

Again it is my contention that Brouwer's claims do not stand up to scrutiny, relative to his own ontological ideas. It is difficult to reconcile the restriction that phase 2 has to be after phase 1 with the restriction that phase 3 has to be after phase 2. To uphold the former restriction, one would presumably have to claim something along the lines that a property is abstracted from some objects xx of which it holds and that its identity is dependent on it being abstracted from exactly xx and not some other objects (otherwise a property abstracted from some objects created prior to it would also be "supposable for" further objects created later). But then the property already holds for the objects xx at stage 2, in conflict with the latter restriction.

Unless a property is identified with its extension, and Brouwer certainly doesn't do that, there is no clear reason to think that it can't exist prior to or independently of the objects of which it holds. It would seem that it can according to the intuitionist ontology, where the property is a separate construct. And it can as well in the non-verificationist constructivism developed here, for we can identify a class (as I choose to call it) with a linguistic rule. As addition

[4]The intuitionist position is stated more clearly by Heyting (1931, 111): "Eine Spezies wird, ebensowenig wie eine Menge, als Inbegriff ihrer Elemente betrachtet, sondern mit ihrer definierenden Eigenschaft identifiziert. Imprädicative Definitionen sind schon hierdurch unmöglich, daß, wie für den Intuitionisten von selbst spricht, als Elemente einer Spezies nur vorher definierte Gegenstände auftreten können." ("A species, like a spread, is not regarded as the sum of its members but is rather identified with its defining property. Impredicative definitions are made impossible by the fact, which intuitionists consider self-evident, that only previously defined objects may occur as members of species.")

[5]I am following the interpretation in (van Stigt 1990, 338). The reason the characterization is loose is that the number 54.369 does not have to have been individually constructed along with each of the other natural numbers in phase 1 for the species of natural numbers to be created in phase 2. Subtleties about what *could have been* created prior to the species and about being *intensionally* completed as opposed to extensionally completed are necessary for a full understanding of Brouwer, but are not relevant for present purposes.

is defined by two rules for when a sentence can be made true based on other sentences that are already true, a class can be defined as a rule governing when a sentence saying that some object is an element of the class can be made true based on some other sentence being true. As such a linguistic rule can exist independently of its applications and the truths to which it can be applied, there is no basis for demanding a hierarchy.

We will use the notation "$\{x|\phi\}$" for the class of objects x that satisfy ϕ, where ϕ is a formula with at most the variable x free. The rule then is that a sentence of the form $\psi \in \{x|\phi\}$ can be made true (false) when the sentence $\phi(x/\psi)$ is already true (false). Identifying the class with the linguistic expression "$\{x|\phi\}$" and this rule, the class can exist prior to the TIC and FIC of any sentence of the form $\psi \in \{x|\phi\}$. It does when it has been formulated but the rule has not yet been applied in any specific case. There is therefore no obstacle to a class becoming a member of itself. This just requires a process starting with the formulation of a class "$\{x|\phi\}$" and continuing with a series of sentences being made TIC or FIC, culminating in the sentence $\phi(x/\{x/\phi\})$ and then the sentence $\{x|\phi\} \in \{x|\phi\}$ being made TIC.

Here is a simple example. Let ϕ_1 be the predicate of being an even number, i.e. $\exists m(2 \cdot m = n)$, while ϕ_2 is the predicate of being a non-empty class, formalized $\exists y(y \in x)$. First the sentence $2 \cdot 3 = 6$ is made true as explained in the previous chapter. Then, by one of the two rules for the existential quantifier in the previous chapter, the sentence $\exists m(2 \cdot m = 6)$ is made true. The latter is $\phi_1(n/6)$, so after that, the sentence $6 \in \{n|\phi_1\}$ can be made true using the class rule. Employing the existential rule again, the sentence $\exists y(y \in \{n|\phi_1\})$ can be made true. This sentence is again identical to $\phi_2(x/\{n|\phi_1\})$, ergo $\{n|\phi_1\} \in \{x|\phi_2\}$ can be made true. Using a version of the existential rule with classes in place of natural numbers, we can make $\exists y(y \in \{x|\phi_2\})$, which is identical to $\phi_2(x/\{x/\phi_2\})$, true. Then it is possible to make the class of non-empty classes a member of itself: $\{x|\phi_2\} \in \{x|\phi_2\}$.

When we conceive of classes this way, it is thus legitimate to employ impredicative definitions like "the class of all those classes that...". We are not in conflict with the vicious circle principle, which in Gödel's (1983, 454) formulation reads "no totality can contain members definable only in terms of this totality", for a class is not an extensional "totality" but an intensional criterion of membership. For the same reason, the membership relation does not have to be well-founded, even though the dependency relation between truths about membership does.

Language playing a constitutive role, meaning that a class is not primarily an extensional collection, but a linguistic entity, identified with its name plus its criterion of membership, has the consequence that identity of classes is not a matter of coextensionality (although a relation of equality can be defined from that) but is an intensional relation: two classes are identical if they have the same membership criterion. Another consequence is that there are no classes except those that can be described in language.

To sum up: I agree with Brouwer that we should work with a notion of collection that is distinguished by criteria of membership, rather than being combinatorial in nature. I disagree with him on two important points. My concept of classes will allow for circularity, while Brouwer's species are typed. And where membership for Brouwer is a matter of embeddability of one non-linguistic construction into another, it will, in the class theory to be developed, be a matter of sentences asserting membership being made true on the basis of other sentences being true, similarly to the rule-based arithmetic of Section 4.4.

5.2 Kripke's theory of truth

A formal theory of classes that codifies the above example already exists. It was formulated by Maddy in her two papers (1983) and (2000), building on Kripke's (1975) theory of truth. The plan for the rest of this chapter is to introduce Kripke's theory, introduce a theory that resembles Maddy's, expand it into a theory that also covers real numbers, discuss some problems that they both have, and study diagonalization in this setting. Dealing with the problems will be the aim of Chapters 6 and 7.

More formal precision will be needed in this chapter than earlier in the dissertation, so let us carefully specify Kripke's theory. It is given in classical set theory. We begin with the syntax of the language: For each $n \in \mathbb{N}$, there is a countable set $\mathcal{P}_n$ of **ordinary** n**-ary predicates**. In addition there is the **truth predicate** T. We also have a countable infinite set $\mathcal{C}$ of **constants** and a countable set of **variables**. Variables and constants are called **terms**.

The set of **formulae** is defined recursively:

- If P is an n-ary predicate (ordinary or the truth predicate) and $t_1, \dots, t_n$ are terms, then $P(t_1, \dots, t_n)$ is a formula.
- If ϕ and ψ are formulae, then $\neg\phi$ and $\phi \vee \psi$[6] are formulae.
- If ϕ is a formula and x a variable, then $\exists x \phi$ is a formula.
- Nothing is a formula except by virtue of the above clauses.

[6]To be precise this should be "$(\phi \vee \psi)$", but to improve on readability I will be systematically sloppy with brackets.

Conjunction, the conditional, the bi-conditional and universal quantification are defined as abbreviations in the usual way. A formula is a **sentence** if it does not contain any free variables. Let $\mathcal{S}$ be the set of sentences.

A **model** is a pair $\mathfrak{M} = (D, I)$ such that

- D, the **domain**, is a superset of $\mathcal{S}$,
- I, the **interpretation function**, is a function defined on $\bigcup_{n\in\mathbb{N}} \mathcal{P}_n \cup \mathcal{C}$ such that
  - for every $P \in \mathcal{P}_n$, $I(P) \subseteq D^n$,
  - for every $c \in \mathcal{C}$, $I(c) \in D$, and
  - $\{I(c) | c \in \mathcal{C}\} = D$.

We dispense with the technicalities of Gödel numbers and arithmetisation by facilitating self-reference simply by letting the domain include all sentences of the language and making assumptions about the model when needed, e.g. that for some constant c_l it holds that $I(c_l) = \neg T(c_l)$, i.e. that there is a liar sentence.[7]

In accordance with the thesis from the last chapter that it is possible to have names for everything, it has been assumed that there is a constant for each object in the domain. Thus the semantics of the quantifier can be defined substitutionally.

The predicate T is interpreted by an **evaluation** $\mathcal{E} = (T, F)$ where T and F are subsets of $\mathcal{S}$, called respectively the **truth set of** $\mathcal{E}$ and the **falsity set of** $\mathcal{E}$.[8] If a sentence ϕ is in T (F) we can express this by saying that ϕ **is true (false) in** $\mathcal{E}$. We say that $\mathcal{E}$ is **consistent** if T and F are disjoint. We also say that an evaluation $\mathcal{E}' = (T', F')$ **extends** $\mathcal{E}$ if $T \subseteq T'$ and $F \subseteq F'$.

The **evaluation with respect to (the model $\mathfrak{M}$ and) the evaluation** $\mathcal{E} = (T, F)$, in symbols $\mathrm{E}_{\mathcal{E}}$,[9] is defined as $(T_{\mathcal{E}}, F_{\mathcal{E}})$, where $T_{\mathcal{E}}$ and $F_{\mathcal{E}}$ are defined recursively by the following clauses:

1) If ξ is of the form $P(c_1, \ldots, c_n)$ where P is an ordinary n-ary predicate and $c_1, \ldots, c_n$ are constants, then
   - $\xi \in T_{\mathcal{E}}$ if $(I(c_1), \ldots, I(c_n)) \in I(P)$, and

[7] I'm here following (Gupta 1982).

[8] The letter "T" is thus used both for the truth predicate in the object language and the set of true sentences in the meta-language. This should not cause confusion.

[9] Since the evaluation is relativised to a model, it would be more correct to use the notation "$\mathrm{E}_{\mathfrak{M},\mathcal{E}}$" instead of "$\mathrm{E}_{\mathcal{E}}$". But we will never consider more than one specific model at a time, and hence the subscript indicating the model can be dispensed with in the interest of simple notation. For the same reason, universal quantification over models will be implicit in much of what follows.

$\neg\phi$		
	$\top$	$\bot$
ϕ	$\bot$	$\top$
	$+$	$+$

$\phi \vee \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\top$	$\top$
ϕ	$\bot$	$\top$	$\bot$	$+$
	$+$	$\top$	$+$	$+$

$\phi \wedge \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\bot$	$+$
ϕ	$\bot$	$\bot$	$\bot$	$\bot$
	$+$	$+$	$\bot$	$+$

$\phi \rightarrow \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\bot$	$+$
ϕ	$\bot$	$\top$	$\top$	$\top$
	$+$	$\top$	$+$	$+$

$\phi \leftrightarrow \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\bot$	$+$
ϕ	$\bot$	$\bot$	$\top$	$+$
	$+$	$+$	$+$	$+$

Figure 5.1: Strong Kleene truth tables

- $\xi \in F_{\mathcal{E}}$ if $(I(c_1), \ldots, I(c_n)) \notin I(P)$.

2) If ξ is of the form $\neg\phi$ where ϕ is a sentence, then
   - $\xi \in T_{\mathcal{E}}$ if $\phi \in F_{\mathcal{E}}$, and
   - $\xi \in F_{\mathcal{E}}$ if $\phi \in T_{\mathcal{E}}$.

3) If ξ is of the form $\phi \vee \psi$ where ϕ and ψ are sentences, then
   - $\xi \in T_{\mathcal{E}}$ if $\phi \in T_{\mathcal{E}}$ or $\psi \in T_{\mathcal{E}}$, and
   - $\xi \in F_{\mathcal{E}}$ if $\phi \in F_{\mathcal{E}}$ and $\psi \in F_{\mathcal{E}}$.

4) If ξ is of the form $\exists x \phi$ where x a variable and ϕ is a wff with at most x free, then
   - $\xi \in T_{\mathcal{E}}$ if there exists a $c \in \mathcal{C}$ such that $\phi(x/c) \in T_{\mathcal{E}}$, and
   - $\xi \in F_{\mathcal{E}}$ if for all $c \in \mathcal{C}$, $\phi(x/c) \in F_{\mathcal{E}}$.

5) If ξ is of the form $T(c)$ where c is a constant, then
   - $\xi \in T_{\mathcal{E}}$ if there is a sentence ξ' such that $I(c) = \xi'$ and $\xi' \in T$,
   - $\xi \in F_{\mathcal{E}}$ if there is a sentence ξ' such that $I(c) = \xi'$ and $\xi' \in F$, and
   - $\xi \in F_{\mathcal{E}}$ if there is no sentence ξ' such that $I(c) = \xi'$.[10]

For ease of reference, the semantics of the connectives (primitive as well as defined) is put in schematic form in Figure 5.1. Here, the symbols $\top$, $\bot$, and $+$ are used for true, false, and undefined respectively.

$\mathrm{E}_{\mathcal{E}}$ is the "jump" function applied to the evaluation $\mathcal{E}$. Beginning from the empty evaluation, iterating this jump function and collecting up at limit ordinals, we get the Kripke hierarchy: for every ordinal α the **evaluation at**

[10] This last bullet point says that if c does not denote a sentence but something else, then a sentence saying, that the object denoted by c is true, is false. Alternatively it could be left undefined; nothing important hinges on it.

level α, E^α, is defined like this:

$$\mathrm{E}^\alpha = \begin{cases} (\varnothing, \varnothing) & \text{if } \alpha = 0 \\ \mathrm{E}_{\mathrm{E}^{\alpha-1}} & \text{if } \alpha \text{ is a successor ordinal} \\ \left(\bigcup_{\eta<\alpha} T_{\mathrm{E}^\eta}, \bigcup_{\eta<\alpha} F_{\mathrm{E}^\eta}\right) & \text{if } \alpha \text{ is a limit ordinal} \neq 0 \end{cases}$$

The sequence of evaluations is monotonic in the sense that it holds for all ordinals α and β that if $\alpha < \beta$ then E^β extends E^α (Kripke 1975, 703). It reaches a fixed point where no further sentences are added to the truth set or the falsity set: there is an ordinal α such that $\mathrm{E}^{\alpha+1} = \mathrm{E}^\alpha$ – and consequently, for all larger ordinals β, $\mathrm{E}^\beta = \mathrm{E}^\alpha$ (Kripke 1975, 704–705). This is the final interpretation of T, and it is consistent. According to Kripke's theory, a sentence is true if it is in the truth set of this evaluation and false if it is in the falsity set. If a sentence is in neither, we say that its truth value is undefined, or in short that the sentence is undefined.

We will also say that **a sentence is true (false; undefined) at level** α if it is in the truth set (falsity set; neither) of E^α, and that it is **made true (false)** or **becomes true (false) at level** α if it is true (false) at level α and not at any lower level.

The theory as presented here is the strong Kleene, minimum fixed point version of Kripke's theory. In this chapter, where we will not be concerned with other versions, it will simply be referred to it as "Kripke's theory". In the following chapters, where we will, it will be called the "basic version (of Kripke's theory)".

Kripke can celebrate two great victories compared to the Tarskian theory with which he is primarily contrasting his own work. The first is that the theory validates the Tarskian T-schema inside the object language: $T(c)$ is true if and only if $I(c)$ is true (Kripke 1975, 702, 705).[11] The second is that the sentences of the Watergate example (Kripke 1975, 691) and similar examples get the intuitively correct truth value. One version of the example is as follows. Jones asserts exactly one sentence about Watergate, namely

J1: Nixon has spoken falsely about Watergate

while Nixon asserts

N1: Everything Jones says about Watergate is true

[11] Recall the distinction made in Section 4.1 between 1) The sentence "'the star has five vertices' is true", 2) the sentence "the star has five vertices" and 3) the state of affairs that the star has five vertices. It was argued that the relationship that was relevant for the question of the learnability of truth-conditional semantics was that between 2) and 3). In contrast, we are here talking about the relationship between 1) and 2).

and

N2: My administration is not involved in the break in at Watergate

and nothing else on the subject. It is not possible for all of these sentences to be reconstructed in languages of the Tarski hierarchy.[12] For N1 predicates truth of J1 and must therefore be formulated in a language that, relative to the language in which J1 is formulated, is on a meta-level. But conversely, the truth predicate (prefixed by a negation) must be applied to a name of N1 in the formulation of J1, so J1 must be formulated in a meta-language to the language of N1. But intuitively this system of sentences is not pathological. Intuitively, J1 and N1 are true because N2 is false: the falsity of N2 is sufficient to make J1 true and that is again sufficient to make N1 true.

Kripke's theory is in line with this intuitive verdict. The sentences can be formalised as

(J1) $$\exists x(N(x) \wedge \neg T(x)),$$

(N1) $$\forall x(J(x) \rightarrow T(x))$$

and

(N2) $$\neg W(a),$$

where N, J and W are unary predicates meaning "is an utterance about Watergate made by Nixon", "is an utterance about Watergate made by Jones" and "is involved in the break in at Watergate" respectively, and a is a constant denoting Nixon's administration. We therefore have $I(J) = \{(\mathrm{J1})\}$, $I(N) = \{(\mathrm{N1}),(\mathrm{N2})\}$ and $a \in I(W)$. (N2) is made false at level 1. If n_2 is a constant denoting this sentence, then $T(n_2)$ becomes false at level 2 and hence at the same level $\neg T(n_2)$, $N(n_2) \wedge \neg T(n_2)$ and (J1) are made true. Then at level 3, this makes all sentences of the form $T(j_1)$, where j_1 is a constant referring to (J1), true. So at level 3, $J(j_1) \rightarrow T(j_1)$ also becomes true, and for any constant c not referring to (J1), $J(c) \rightarrow T(c)$ was made true already at level 1 by the falsity of the antecedent. Ergo, (N1) also becomes true at level 3.

However, Kripke's theory suffers from the problem that correct generalizations about the semantics tend to come out undefined when formulated in the object language itself. A significant example is that even though the Tarskian T-schema is validated in the sense explained above, the object language sentence that expresses that fact is undefined. The sentence is $\forall x(P(x) \rightarrow T(x))$ where

[12] See (Tarski 1933) and (Tarski 1944).

P is a unary predicate, the interpretation of which is the set of all sentences of the form $T(c_\phi) \leftrightarrow \phi$, where again c_ϕ is a constant denoting the sentence ϕ. There are two reasons for this failure, each individually sufficient. The first is that there are undefined instances of $T(c_\phi) \leftrightarrow \phi$, namely those where ϕ is undefined. The second (which is a special case of the first, but worth singling out) is that the truth value of $\forall x(P(x) \to T(x))$ depends on itself: the sentence is itself one of the sentences that can take the place of ϕ in $T(c_\phi) \leftrightarrow \phi$, thus for it to become true at one of Kripke's levels, it would have to be true or false at a lower level. The same problem (although of course with different specific sentences) exists for the theory of classes to be formulated in this chapter. It will be addressed in the final two chapters. Further problems of expressibility will be mentioned in Chapter 6.

Another problem is that the theory relies on the classical theory of transfinite ordinals and, more generally, classical set theory, which we found reasons to distrust in Chapter 1. This problem will be solved in Chapter 7, where a modified theory of classes will be developed that is in line with the TIC/TAP doctrine of truth. For now we will pretend that the levels make sense as instances of time, in order to focus on other things first. We will both pretend that the transfinite levels can be understood as instances of time that come after an infinity of other instances of time and that infinitely many sentences can be made true or false at one instant of time, for that is what happens at each level.

5.3 A theory of classes and real numbers

In parallel to Kripke's theory, we will formulate a theory of classes where the distinctive rule is that when $\phi(v/\psi)$ is true (false) at some level, $\psi \in \{v|\phi\}$ will be true at the next level. We will use it to say something about real numbers, real numbers being defined as a certain kind of classes, so that we have an alternative to Brouwer's free choice sequence-based account that was discussed and rejected in Chapter 2. In that endeavor we will take quite a bit of inspiration from Bishop (1967).[13] Bishop defines a real number as a sequence $\{x_n\}$ of rational numbers such that $|x_m - x_n| \leq m^{-1} + n^{-1}$ for all natural numbers n and m (p. 15), and a sequence as "a rule which associates to each positive integer n a mathematical object a_n" (p. 12), and so will we.

We thus need natural numbers and (more generally) rational numbers. We could take the rules constituting arithmetic from the previous chapter and

[13]Bishop developed a mathematics that can be described as the intersection of classical mathematics and intuitionism. We will not be concerned with his theory as such, and no exegetical claims will be made, we will just take the bits we can use.

incorporate them into the formal theory for classes. However, formulating the theory will be complicated enough even without, so instead we just import arithmetic in a "black box" in the form of an interpretation function. Further, I trust that the reader will grant me that having vindicated classical arithmetic, the classical theory of rational numbers could also be recovered, and we will therefore do the same with rationals.

We will make use of a multi-sorted language. The syntax is specified as follows. First, by way of constants, we have one numeral, called a **rational numeral**, for each rational number. A rational numeral is also called a **natural numeral** when it is for a natural number (not including 0). Second, there are three disjoint, countable sets of **natural variables**, **rational variables** and **class variables**. The intention is that the natural variables range over the natural numerals, the rational variables range over the rational numerals and the class variables range over the classes (to be defined), but that is made precise below.

For numerals and variables we use these notational conventions: n, m and l with or without some subscript are natural variables; q, x and y with or without some subscript are rational variables; and a, b, c, d and σ with or without some subscript are class variables. For specific numerals we simply use the normal arabic numerals or the normal fraction notation, while capital letters are used as meta-language variables ranging over object language numerals and classes. In that case the same notational conventions regarding which letters are used for natural and rational numerals and classes apply. Subscripts are used purely as a mnemonic device. So for instance "x_{n-1}" is meant as an unanalyzable symbol, i.e. "n" is not itself a variable into whose position a constant can be substituted.

We also have countably many **predicates**. For present purposes, predicates that apply to classes (and class variables), and only such as are unary, will do.

We define the set of **terms** to be the smallest set which includes the rational numerals and the rational variables and also contains the following whenever $t_1, \ldots, t_n$ are terms and t is a natural numeral: $t_1 + t_2$, $t_1 - t_2$, $t_1 \cdot t_2$, t_1/t, t_1^t, t_1^{-t}, $\min\{t_1, \ldots, t_n\}$, $\max\{t_1, \ldots, t_n\}$ and $|t_1|$. The set of **natural terms** is the smallest set containing the natural numerals and the natural variables and containing $t + t'$ and $t \cdot t'$ whenever t and t' are natural terms.

The set of **formulae** and the set of **classes**[14] are defined by simultaneous recursion, like this:

[14] They are called "classes" instead of "class terms" because the latter would give the impression that the linguistic expression refers to something different from itself. But the point is that the class just is the linguistic expression together with the rules that govern it (to follow).

- If t_1 and t_2 are terms, then $t_1 < t_2$, $t_1 \leq t_2$ and $t_1 \equiv t_2$ are formulae.
- If t_1 is a term and t_2 is a class or class variable, then $t_1 \in t_2$ is a formula.
- If t_1 and t_2 are each either a class or a class variable, then $t_1 \equiv t_2$ and $t_1 \in t_2$ are formulae.
- If P is a predicate and t_1 is a class or a class variable, then $P(t_1)$ is a formula.
- If ϕ and ψ are formulae, then $\neg\phi$ and $\phi \vee \psi$ are formulae.
- If ϕ is a formula and v a variable (of any kind), then $\exists v\phi$ is a formula.
- If ϕ is a formula and v a variable (of any kind), then $\{v|\phi\}$ is a class (ϕ is then called the **defining formula** of the class).
- Nothing is a formula or a class except by virtue of the above clauses.

A notion of complexity of formulae will be needed. A rigorous definition can be dispensed with in favor of these stipulations: A formula constructed according to one of the first four bullet points has minimal complexity (even if built using complex classes). And if it is bullet five or six that has been used in "the last step" of the construction, the formula in question has higher complexity than both ϕ and (in the case of bullet five) ψ.

The formulas $t_1 > t_2$ and $t_1 \geq t_2$ of course mean the same as $t_2 < t_1$ and $t_2 \leq t_1$ respectively. The other connectives and the universal quantifier are again defined in the usual way. We also define the symbol $\notin$, restricted quantification and the notation for unique existence ($\exists! v\phi$) as usual.

Free and bound occurrences of variables are defined as usual, we just need to stipulate that v is bound in $\{v|\phi\}$. A formula is called a **sentence** and a term or a class is called **closed** if it does not contain any free variables (in the case of terms this simply means that it contains no variables at all, as there is nothing in a term itself that can bind a variable). Finally, again let $\mathcal{S}$ be the set of sentences.

To specify the semantics, we start out with an **interpretation function**, I, with the union of the set of closed terms and the set of predicates as domain. It interprets closed terms as rational numbers. It is defined recursively in the obvious way. So something like $|19 + 2^{-7}|$ means exactly what it looks like. The interpretation function interprets predicates as subsets of the set of classes.

We want the relation symbol "$\equiv$" to express intensional identity. Between closed terms that is accomplished straightforwardly by letting a sentence claiming two terms to be related by $\equiv$ to be true if the two relata are given the same value by the interpretation function and false otherwise.

Having "$\equiv$" express intensional identity between classes is a little more complicated.[15] This is, intuitively, a more inclusive relation than syntactic identity, for e.g. the class $\{n_1|\exists n_2(n_2 \cdot 2 \equiv n_1)\}$ and the class $\{n_1|\exists n_3(\frac{4}{2} \cdot n_3 \equiv n_1)\}$ clearly mean the same thing, namely "the class of even numbers". We capture that intuition in a definition of synonymy between classes, and in the formulation of that definition we employ an auxiliary notion of synonymy between terms:

Two terms t and t' are **synonymous** if $n_1, \ldots, n_i, q_1, \ldots, q_j$ are the variables which occur in either one of them and for any choice of constants $N_1, \ldots, N_i, Q_1, \ldots, Q_j$, it is the case that

$$I(t(n_1/N_1)\ldots(n_i/N_i)(q_1/Q_1)\ldots(q_j/Q_j))$$

is identical to

$$I(t'(n_1/N_1)\ldots(n_i/N_i)(q_1/Q_1)\ldots(q_j/Q_j)).$$

Two closed classes are **synonymous** if they are related by the transitive and reflexive closure of the (obviously symmetrical) relation defined by stipulating that C is related to C' if one of these conditions is satisfied: 1) C' is the result of substituting an occurrence of a term t in C with a synonymous term t'. 2) C' is the result of substituting all occurrences of a variable v in C which are bound by the same quantifier or class abstractor (i.e. in the symbol combinations "$\exists v$" and "$\{v|$") with a variable v' of the same kind (natural, rational or class) such that no substituted occurrence of v' is bound by some quantifier or class abstractor which the corresponding occurrence of v was not bound by.

So returning to the example above, $\{n_1|\exists n_2(n_2 \cdot 2 \equiv n_1)\}$ is synonymous with $\{n_1|\exists n_3(\frac{4}{2} \cdot n_3 \equiv n_1)\}$ according to this definition, because the class $\{n_1|\exists n_2(\frac{4}{2} \cdot n_2 \equiv n_1)\}$ results from the former by substitution of a synonymous term (for any constant that is substituted into the position of n_2 in $\frac{4}{2} \cdot n_2$ and $n_2 \cdot 2$, the interpretation function gives the same value when applied to the two results) and the latter results from this "intermediate class" by uniform substitution of a variable.[16]

[15] The symbol "$\equiv$" is used because extensional identity between classes, for which the symbol "$=$" would be proper, *could* be added by definition. In that case it would be a relation whose extension would be determined only gradually in the iterative process, unlike the intensional identity which is fixed from the outset. In other words, sentences of the form $C \equiv C'$ are all made true or false at level 1, while sentences of the form $C = C'$ could be given a truth value at any level or remain undetermined through all the levels. (In fact, it is a more precise way to characterize $\equiv$ to say that it identifies what can be identified at level 0, than to say that it is intensional.)

[16] The definition of synonymity is good enough for present purposes and its not aiming at being more than that. For instance, $\{n_1|\exists n_2(n_2 \cdot 2 \equiv n_1)\}$ is not synonymous with $\{n_1|\exists n_2(n_1 \equiv n_2 \cdot 2)\}$ according to this definition.

The definitions of an evaluation, an evaluation being consistent and one evaluation extending another are as before.

The Kripkean "jump" is specified by defining the **evaluation with respect to the evaluation** $\mathcal{E} = (T, F)$, $\mathrm{E}_{\mathcal{E}}$, as $(T_{\mathcal{E}}, F_{\mathcal{E}})$, where $T_{\mathcal{E}}$ and $F_{\mathcal{E}}$ are defined by recursion on the complexity of the sentence as follows:[17]

E1) If ξ is of the form $t < t'$ or $t \leq t'$, respectively, where t and t' are closed terms, then
- $\xi \in T_{\mathcal{E}}$[18] if $I(t)$ is less than $I(t')$ or $I(t)$ is less than or equal to $I(t')$, respectively, and
- $\xi \in F_{\mathcal{E}}$ otherwise.

E2) If ξ is of the form $t \equiv t'$ where t and t' are closed terms, then
- $\xi \in T_{\mathcal{E}}$ if $I(t)$ and $I(t')$ are identical, and
- $\xi \in F_{\mathcal{E}}$ otherwise.

E3) If ξ is of the form $s \equiv s'$ where s and s' are closed classes, then
- $\xi \in T_{\mathcal{E}}$ if s and s' are synonymous, and
- $\xi \in F_{\mathcal{E}}$ otherwise.

E4) If ξ is of the form $t \in \{n|\phi\}$ where t is a closed term and $\{n|\phi\}$ is a closed class, then
- $\xi \in T_{\mathcal{E}}$ if t is a natural term and $\phi(n/t) \in T$, and
- $\xi \in F_{\mathcal{E}}$ if t is not a natural term or $\phi(n/t) \in F$.

E5) If ξ is of the form $t \in \{q|\phi\}$ where t is a closed term and $\{q|\phi\}$ is a closed class, then
- $\xi \in T_{\mathcal{E}}$ if $\phi(q/t) \in T$, and
- $\xi \in F_{\mathcal{E}}$ if $\phi(q/t) \in F$.

E6) If ξ is of the form $t \in \{c|\phi\}$ where t is a closed term or a closed class and $\{c|\phi\}$ is a closed class, then
- $\xi \in T_{\mathcal{E}}$ if t is a class and $\phi(c/t) \in T$, and
- $\xi \in F_{\mathcal{E}}$ if t is a term or $\phi(c/t) \in F$.

E7) If ξ is of the form $P(t)$ where P is a predicate and t is a closed class, then

[17] Clauses E8–E12 make reference to $T_{\mathcal{E}}$ and $F_{\mathcal{E}}$, but only with respect to less complex sentences than the one under consideration. By clauses E4–E6 a sentence may "gain" its truth value from a more complex sentence, but here it is only the truth value as given by $\mathcal{E} = (T, F)$ that is referred to. In short, the truth value of a sentence only depends on the previous level and sentences of lower complexity. Hence, as stated, the definition is simply by recursion on the complexity of the sentence.

[18] In this list the symbol "$\in$" is used both in the object language and the meta-language but this should cause no confusion.

- $\xi \in T_{\mathcal{E}}$ if $I(c) \in I(P)$, and
- $\xi \in F_{\mathcal{E}}$ if $I(c) \notin I(P)$.

E8) If s is of the form $\neg\phi$ where ϕ is a sentence, then
- $s \in T_{\mathcal{E}}$ if $\phi \in F_{\mathcal{E}}$, and
- $s \in F_{\mathcal{E}}$ if $\phi \in T_{\mathcal{E}}$.

E9) If ξ is of the form $\phi \vee \psi$ where ϕ and ψ are sentences, then
- $\xi \in T_{\mathcal{E}}$ if $\phi \in T_{\mathcal{E}}$ or $\psi \in T_{\mathcal{E}}$, and
- $\xi \in F_{\mathcal{E}}$ if $\phi \in F_{\mathcal{E}}$ and $\psi \in F_{\mathcal{E}}$.

E10) If ξ is of the form $\exists n\phi$ where ϕ is a formula with at most n free, then
- $\xi \in T_{\mathcal{E}}$ if there exists an N such that $\phi(n/N) \in T_{\mathcal{E}}$, and
- $\xi \in F_{\mathcal{E}}$ if for all N, $\phi(n/N) \in F_{\mathcal{E}}$.

E11) If ξ is of the form $\exists q\phi$ where ϕ is a formula with at most q free, then
- $\xi \in T_{\mathcal{E}}$ if there exists a Q such that $\phi(q/Q) \in T_{\mathcal{E}}$, and
- $\xi \in F_{\mathcal{E}}$ if for all Q, $\phi(q/Q) \in F_{\mathcal{E}}$.

E12) If ξ is of the form $\exists c\phi$ where ϕ is a formula with at most c free, then
- $\xi \in T_{\mathcal{E}}$ if there exists a C such that $\phi(c/C) \in T_{\mathcal{E}}$, and
- $\xi \in F_{\mathcal{E}}$ if for all C, $\phi(c/C) \in F_{\mathcal{E}}$.

Relative to this new definition, the **evaluation at level** α, E^{α}, is defined as in Kripke's theory. The definitions of a sentence being true in an evaluation, becoming true at a level, etc. also carries over from Section 5.2.

We can then go on to prove that this sequence of evaluations, like Kripke's, is monotonically increasing, that it consists solely of consistent evaluations and that it reaches a fixed point.

Lemma 5.1. *For all ordinals α and β such that α is smaller than β, the evaluation E^{β} extends E^{α}.*

Proof. This is proved by complete induction on β. The base case is vacuous, for if β is 0 then there are no smaller ordinals, and the limit case is trivial. So assume that β is a successor ordinal. We prove that E^{β} is an extension of $\mathrm{E}^{\beta-1}$ from which the conclusion follows from the induction hypothesis. Assume that ξ is true in $\mathrm{E}^{\beta-1}$. We then need to show that ξ is true in E^{β} (the case of falsity is similar), and this is done by inner induction on the complexity of ξ.

If ξ is of the form $t < t'$ where t and t' are closed terms, then it follows from E1 that $I(t)$ is smaller than $I(t')$ so ξ is true in E^{β}. The argument is analogous for the other forms mentioned in E1–E3 and E7.

If ξ is of the form $t \in \{n|\phi\}$ where t is a closed natural term and $\{n|\phi\}$ is a closed class then by E4, if $\beta - 1$ is also a successor ordinal, $\phi(n/t)$ is true in $\mathrm{E}^{\beta-2}$, so by the outer induction hypothesis, $\phi(n/t)$ is also true in $\mathrm{E}^{\beta-1}$, and therefore $t \in \{n|\phi\}$ is true in E^{α}. If instead $\beta - 1$ is a limit ordinal then either it is equal to 0, in which case the conclusion is trivial, or there is an ordinal γ which is smaller and a successor, such that $t \in \{n|\phi\}$ is true in E^{γ}. Then it follows that $\phi(n/t)$ is true in $\mathrm{E}^{\gamma-1}$ and hence – again by the outer induction hypothesis – in $\mathrm{E}^{\beta-1}$, so $t \in \{n|\phi\}$ is true in E^{β}. This type of argument works for the forms mentioned in E5 and E6 as well, so the base case is covered.

The induction step in the inner induction consists of the forms mentioned in E8–E12 of which we just treat E9 explicitly. So assume that ξ is of the form $\phi \vee \psi$ where ϕ and ψ are sentences. Then by E8, either ϕ is true in $\mathrm{E}^{\beta-1}$ or ψ is true in $\mathrm{E}^{\beta-1}$, so it follows using the inner induction hypothesis that either ϕ is true in E^{β} or ψ is true in E^{β}, and ergo $\phi \vee \psi$ is true in E^{β}. □

Theorem 5.2. *For every ordinal α, the evaluation E^{α} is consistent.*

Proof. The structure of this proof is the same as in the previous: outer complete induction on α and, for the successor case, inner induction on complexity of formula. The case where α is 0 is trivial. And the limit case is easiest done by *reductio*: Assume that E^{α} is inconsistent and let ϕ be a sentence which is both true and false in E^{α}. Then there are ordinals β and β' smaller than α such that ϕ is true in E^{β} and false in $\mathrm{E}^{\beta'}$. Let β'' be the largest of β and β'. Then it follows from the Lemma that ϕ is both true and false in $\mathrm{E}^{\beta''}$, making $\mathrm{E}^{\beta''}$ inconsistent. This contradicts the induction hypothesis.

Now let α be a successor ordinal and ϕ a sentence. If ϕ is of one of the forms mentioned in E1–E3 or E7, it is obvious that ϕ can not both be true and false in E^{α}, for the criteria for being made true and false in these clauses are contradictory. In each of E4–E6 it is seen that for ϕ to be both true and false in E^{α} there would have to be some sentence ψ which was both true and false in $\mathrm{E}^{\alpha-1}$, contradicting the outer induction hypothesis. And in E8–E12, ϕ could only be both true and false in E^{α} if some sentence ψ of lower complexity was also both true and false in E^{α}, contradicting the inner induction hypothesis. □

Theorem 5.3. *A unique evaluation $\mathcal{E}$ exists such that for some ordinal α, E^{α} is identical to $\mathcal{E}$, and for all ordinals β larger than α, E^{β} is also identical to $\mathcal{E}$.*

Proof. Uniqueness follows trivially from existence, and existence follows from monotonicity together with the fact that there are only countably many sentences. For according to the Lemma, the sequence of evaluations is increasing, which means that there cannot be a larger number of different evaluations in the sequence than there are sentences. But it cannot be strictly increasing, i.e. it cannot be the case that for any pair of distinct ordinals α and β, the evaluations E^α and E^β are different, for then there would be uncountably many evaluations (as many as there are ordinals). So for some α, E^α is identical to $\mathrm{E}^{\alpha+1}$. And then we can prove by induction that for any β larger than α, E^β is identical to E^α: If β is a successor ordinal then E^β is identical to $\mathrm{E}_{\mathrm{E}^{\beta-1}}$ which by the induction hypothesis is identical to $\mathrm{E}_{\mathrm{E}^\alpha}$ which equals $\mathrm{E}^{\alpha+1}$ alias E^α. And if β is a limit ordinal, then it follows that the truth set of E^β is a union of sets which are identical to the truth set of E^α (by the induction hypothesis) or subsets thereof (by the Lemma), and similarly for the falsity set, so E^β is identical to E^α. □

We call a sentence **true** (**false**) *simpliciter* if it is true (false) in the $\mathcal{E}$ of the Theorem, and **undefined** if it is neither true nor false.

Notice that the proof of Theorem 5.3 relies on the Cantorian theory of the transfinite which has been rejected. It needs to be shown that the theory can be reconstructed without this reliance. But as already noted, this challenge is postponed.

5.4 Unrestricted comprehension

With the presentation of the formal system completed, we can now go on to investigate its properties. The first subject will be comprehension, for which a simple theorem holds:

Theorem 5.4. *For every formula ϕ with exactly one free variable n and every N, $\phi(n/N)$ is true iff $N \in \{n|\phi\}$ is true. Similarly for rational variables and numerals and for class variables and classes.*

Proof. Assume that $\phi(n/N)$ is true. There is a level at which $\phi(n/N)$ is made true. By E4, $N \in \{n|\phi\}$ is true at the next level, so because of monotonicity it is true in the fixed point. Now assume instead that $N \in \{n|\phi\}$ is true. The level at which it is made true must be a successor level, so at the previous level $\phi(n/N)$ is true. And again the truth of $\phi(n/N)$ in the fixed point follows from Lemma 5.1. □

This result matches the unrestricted validity of the T-schema in Kripke's theory. In that theory the paradoxicality of semantic self-reference is avoided by having the Liar and related sentences undefined. It is similar here when it comes to those "paradoxical" classes that the classical restriction of comprehension is designed to eliminate. We will have a look at Russell's Class and the class of all classes but start out softly with the empty class.

An empty class can be given as $\{c|\neg(c \equiv c)\}$. For any closed class C, $\neg(C \equiv C)$ is made false at level 1 by E3 and E8. So at level 2, $C \in \{c|\neg(c \equiv c)\}$ is made false by E6.

Notice that it is "*an* empty class" and not "*the* empty class". The class $\{c|0 \equiv 1)\}$ is also empty and it is different from $\{c|\neg(c \equiv c)\}$ because they are different as syntactic objects.

In this theory there is no problem in defining the dual of empty classes, a class of all classes: $\{c|c \equiv c\}$. In strict analogy, we have that for any closed class C, $C \equiv C$ is made true at level 1, so that at level 2, $C \in \{c|c \equiv c\}$ is made true. Theorem 5.2 assures us that this does not give rise to a contradiction.

In particular, this class of all classes is an element of itself: $\{c|c \equiv c\} \in \{c|c \equiv c\}$. Self-membership is possible and we can define a class of all classes that are elements of themselves. Or, more interestingly, a class of all classes that are not elements of themselves, i.e. Russell's Class: $\{c|c \notin c\}$. At level 3 it becomes true that $\{c|\neg(c \equiv c)\}$ is an element of Russell's Class and false that $\{c|c \equiv c\}$ is. The sentence saying that Russell's Class is an element of itself is $\{c|c \notin c\} \in \{c|c \notin c\}$. By E6, this sentence depends for its truth value on the sentence $\{c|c \notin c)\} \notin \{c|c \notin c\}$, i.e. the sentence claiming that Russell's Class is *not* an element of itself. And from E8 it is seen that this sentence depends on the original sentence.[19] So the circularity implies (by an induction argument) that neither of the two sentences are ever given a truth value. Therefore, the parallel to Kripke's theory is perfect: as the Liar is meaningful and doesn't lead to inconsistency in Kripke's theory because it is undefined, Russell's Class exists and does not lead to inconsistency in this theory because it is undefined whether it is an element of itself.

Let us return to classes of all classes. There are two reasons why a set of all sets cannot be "allowed" to exists in classical mathematics. The first is that if it does, then it follows from the axiom of restricted comprehension that Russell's Set also exists, which leads to paradox. We have just seen how that problem is defused here. But the second reason is that diagonalization shows that the power set of a set is always larger than that set, so if there is a set of all sets, which by its nature is a superset of its own power set, then it is larger than itself. In the sequel we will define reals and sequences of reals as certain

[19]Here "ϕ depends on ψ" means that ϕ will only become true or false at some level, if ψ has a truth value at the same or the previous level. We will not need a definition of dependency that is adequate for more than just this example, so we shall not try to come up with one. Such definitions are given in (Yablo 1982), (Bolander 2003), (Leitgeb 2005) and (Hansen 2014).

kinds of classes, and for a given sequence of reals define the diagonalization of that sequence, again following Bishop, and we will show that that problem also does not arise in this setting.

A more sustained discussion of failure of bivalence will follow in Chapter 7, in particular a discussion of Russell's Class, henceforth denoted "$\mathcal{R}$".

5.5 The basics of classes and sequences

The definition of diagonalization and the theorem and proof concerning it are quite complex. Therefore, we will build up to them gradually and use the Fibonacci sequence and square roots as examples to introduce some of the techniques in a simpler setting. First, definitions of the notations for ordered pairs and triples and a lemma about them: $\{t', \ldots, t^{(n)}\}$ is defined to be $\{q|q \equiv t' \vee \ldots \vee q \equiv t^{(n)}\}$ when $t', \ldots, t^{(n)}$ are terms and q is a rational variable that does not appear in any of these. If instead $t', \ldots, t^{(n)}$ are classes or class variables, then $t \equiv \{t', \ldots, t^{(n)}\}$ is $\{c|c \equiv t' \vee \ldots \vee c \equiv t^{(n)}\}$ with a similar restriction on which class variable c can be. With that, the ordered pair $\langle t, t' \rangle$ can be introduced as $\{\{t\}, \{t, t'\}\}$, and the ordered triple $\langle t, t', t'' \rangle$ as $\langle \{t\}, \langle t', t'' \rangle \rangle$, when t, t' and t'' are terms.

Lemma 5.5. *Let t_1, t_1', t_2 and t_2' be closed terms. The sentence $\langle t_1, t_1' \rangle \equiv \langle t_2, t_2' \rangle$ is made true or false at level 1; true if $I(t_1)$ equals $I(t_2)$ and $I(t_1')$ equals $I(t_2')$, false otherwise. Similar for $\langle t_1, t_1', t_1'' \rangle \equiv \langle t_2, t_2', t_2'' \rangle$.*

Proof. We just prove the case of $\langle t_1, t_1' \rangle \equiv \langle t_2, t_2' \rangle$. Written out in full that sentence looks like this:

$$\begin{aligned}\{c_1|c_1 \equiv \{q_{1a}|q_{1a} \equiv t_1\} \vee c_1 \equiv \{q_{1b}|q_{1b} \equiv t_1 \vee q_{1b} \equiv t_1'\}\} \equiv \\ \{c_2|c_2 \equiv \{q_{2a}|q_{2a} \equiv t_2\} \vee c_2 \equiv \{q_{2b}|q_{2b} \equiv t_2 \vee q_{2b} \equiv t_2'\}\}\end{aligned}$$

As there are no variables in t_2 and t_2', the right-hand side class is synonymous with the result of replacing q_{2a} with q_{1a}, q_{2b} with q_{1b} and c_2 with c_1. That class is again synonymous with the left-hand side class if and only if $I(t_1)$ equals $I(t_2)$ and $I(t_1')$ equals $I(t_2')$, so in that case the sentence is made true at level 1 by E3 and otherwise false. □

The Fibonacci sequence is $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$ where every number except the two first are the sum of the two previous. We will model a sequence as a rule that describes a class of ordered pairs such that for each natural number there is a unique pair with that number as first element in the class. So the rule for the Fibonacci sequence should describe $\langle 1, 0 \rangle$, $\langle 2, 1 \rangle$, $\langle 3, 1 \rangle$, etc. We can identify the Fibonacci sequence with $\mathcal{F} := \{c_n | \exists\sigma(P(\sigma) \wedge \mathcal{F}')\}$ where we define $\mathcal{F}'$ to be

$$\begin{aligned} &c_n \equiv \langle 1, 0\rangle \vee c_n \equiv \langle 2, 1\rangle \vee \exists c_{n-2} \in \sigma, c_{n-1} \in \sigma, n, m_{n-2}, m_{n-1} \\ &\qquad (c_{n-2} \equiv \langle n-2, m_{n-2}\rangle \wedge c_{n-1} \equiv \langle n-1, m_{n-1}\rangle \wedge c_n \equiv \langle n, m_{n-2} + m_{n-1}\rangle) \end{aligned}$$

and where P is a predicate whose interpretation is the singleton of $\mathcal{F}$.[20]

The sentence $\mathcal{F}'(\sigma/\mathcal{F})(c_n/\langle 1, 0\rangle)$ is a disjunction, one of whose disjuncts is $\langle 1, 0\rangle \equiv \langle 1, 0\rangle$. It is therefore made true by E3 and E9 at level 1. The same holds for $\mathcal{F}'(\sigma/\mathcal{F})(c_n/\langle 2, 1\rangle)$. So $\exists\sigma(P(\sigma) \wedge \mathcal{F}'(c_n/\langle 1, 0\rangle))$ and $\exists\sigma(P(\sigma) \wedge \mathcal{F}'(c_n/\langle 2, 1\rangle))$ are also true at level 1 by virtue of E7, E8, E9[21] and E12. Ergo $\langle 1, 0\rangle \in \mathcal{F}$ and $\langle 2, 1\rangle \in \mathcal{F}$ are made true by E6 at level 2.

Lemma 5.5 tells us that at level 1, the sentence

$$\langle 1, 0\rangle \equiv \langle 3-2, 0\rangle \wedge \langle 2, 1\rangle \equiv \langle 3-1, 1\rangle \wedge \langle 3, 1\rangle \equiv \langle 3, 0+1\rangle$$

is also made true. So by repeated applications of E8, E9, E10 and E12, the sentence

$$\begin{aligned} &\exists c_{n-2} \in \mathcal{F}, c_{n-1} \in \mathcal{F}, n, m_{n-2}, m_{n-1} \\ &\qquad (c_{n-2} \equiv \langle n-2, m_{n-2}\rangle \wedge c_{n-1} \equiv \langle n-1, m_{n-1}\rangle \wedge \langle 3, 1\rangle \equiv \langle n, m_{n-2} + m_{n-1}\rangle) \end{aligned}$$

is made true at level 2. This sentence is one disjunct of $\mathcal{F}'(\sigma/\mathcal{F})(c_n/\langle 3, 1\rangle)$, which is therefore also made true at level 2. So at level 3, $\langle 3, 1\rangle \in \mathcal{F}$ is made true by E6.

In the same way, at every finite level thereafter one more element is added to the sequence. (To prove this, an induction argument is of course required, but we will wait until the proof of Theorem 5.7 to do things so formally correct). At level ω the sequence is complete.

We also need to realize that for any ordered pair which "shouldn't" be in the sequence, the sentence expressing that it is, is actually false. We first see that for any ordered pair of terms $\langle t, t'\rangle$ where t is not a natural term, $\mathcal{F}'(\sigma/\mathcal{F})(c_n/\langle t, t'\rangle)$ is made false at level 1, for it is a disjunction, the first two disjuncts of which are false by Lemma 5.5 and whose third disjunct has an existentially quantified conjunct $\langle t, t'\rangle \equiv \langle n, m_{n-2} + m_{n-1}\rangle$ which is false for any substitution of natural numerals for n, m_{n-2} and m_{n-1}. Hence $\langle t, t'\rangle \in \mathcal{F}$ is made false at level 2. It then follows that $\mathcal{F}'(\sigma/\mathcal{F})(c_n/\langle 1, t\rangle)$ is made false at

[20] That $I(P) = \{\mathcal{F}\}$ is an assumption about the model $\mathfrak{M} = (D, I)$ similar to the assumption about c_l on page 120.

[21] Remember that $P(\sigma) \wedge \mathcal{F}'(c_n/\langle 1, 0\rangle)$ is short for $\neg(\neg P(\sigma) \vee \neg\mathcal{F}'(c_n/\langle 1, 0\rangle))$.

level 2 for any term t which does not designate 0. For the two first disjuncts are clearly false and the third disjunct contains the conjuncts $c_{n-1} \in \mathcal{F}$, $c_{n-1} \equiv \langle n-1, m_{n-1}\rangle$ and $\langle 1, t\rangle \equiv \langle n, m_{n-2} + m_{n-1}\rangle$ which cannot be simultaneously satisfied: n would have to be replaced with 1, so c_{n-1} would have to be replaced with a class synonymous with an ordered pair with 0 as first element and then the first-mentioned conjunct would be false. A similar argument can be made for $\langle 2, t\rangle$, and from there on it is induction.

5.6 Real numbers

We can now go on to define the class of real numbers, the order relation thereon and the class of sequences of real numbers. We do this closely following definitions 1–5 in section 2.2 of (Bishop 1967). $\mathbb{R}$ is defined to be this class:

$$\{a|\forall n_1 \exists! q_1(\langle n_1, q_1\rangle \in a \wedge \forall n_2 \forall q_2(\langle n_2, q_2\rangle \in a \rightarrow |q_2 - q_1| \leq n_2^{-1} + n_1^{-1}))\}$$

Subtraction is defined by letting $a -_{\mathbb{R}} b$ be short for this class

$$\{c_n|\exists n, m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in a \wedge \langle m, y_{2n}\rangle \in b \wedge m \equiv 2 \cdot n \wedge c_n \equiv \langle n, x_{2n} - y_{2n}\rangle)\}$$

The following Lemma says that the result of subtraction is “recognized” by the object language to be a real.

Lemma 5.6. *For all classes A and B such that $A \in \mathbb{R}$ and $B \in \mathbb{R}$ are true, the sentence $A -_{\mathbb{R}} B \in \mathbb{R}$ is also true.*

Proof. Follow the proof of Bishop's (1967, 17) Proposition 2 and add some translation back and forth between object language and meta-language. □

The class of positive real numbers, $\mathbb{R}^+$, is defined to be

$$\{a|a \in \mathbb{R} \wedge \exists n, q(\langle n, q\rangle \in a \wedge q > n^{-1})\},$$

and the class of non-negative real numbers, $\mathbb{R}^{0+}$, is

$$\{a|a \in \mathbb{R} \wedge \forall n, q(\langle n, q\rangle \in a \rightarrow q \geq -n^{-1})\}.$$

The formula $a <_{\mathbb{R}} b$ is short for $b -_{\mathbb{R}} a \in \mathbb{R}^+$, and $a \leq_{\mathbb{R}} b$ is short for $b -_{\mathbb{R}} a \in \mathbb{R}^{0+}$. Formulas where the inequality symbols are reversed or used twice are defined in the obvious way. The formula $a \neq_{\mathbb{R}} b$ stands for $a <_{\mathbb{R}} b \vee b <_{\mathbb{R}} a$. Finally we define the class of sequences of real numbers, $\mathcal{S}_{\mathbb{R}}$, to be

$$\{c|\forall n \exists! a(\langle n, a\rangle \in c) \wedge \forall a(\langle n, a\rangle \in c \rightarrow a \in \mathbb{R})\}.$$

To illustrate how these real numbers work, we will define $\sqrt{2}$ and $\sqrt{3}$ and show that the sentences $\sqrt{2} \in \mathbb{R}$ and $\sqrt{2} <_{\mathbb{R}} \sqrt{3}$ become true. Define $D_{\sqrt{2}}$ to be $\{c_n | \exists \sigma (P(\sigma) \wedge D'_{\sqrt{2}})\}$ where $D'_{\sqrt{2}}$ is

$$\begin{aligned} c_n \equiv \langle 0,1,2\rangle \vee \exists c_{n-1} \in \sigma, n, x_{n-1}, y_{n-1}(c_{n-1} \equiv \langle n-1, x_{n-1}, y_{n-1}\rangle \\ \wedge ((x_{n-1} + 2^{-n})^2 > 2 \rightarrow c_n \equiv \langle n, x_{n-1}, y_{n-1} - 2^{-n}\rangle) \\ \wedge ((x_{n-1} + 2^{-n})^2 \leq 2 \rightarrow c_n \equiv \langle n, x_{n-1} + 2^{-n}, y_{n-1}\rangle)) \end{aligned}$$

and where P is a predicate whose interpretation is the singleton of $D_{\sqrt{2}}$. The real number $\sqrt{2}$ can then be defined to be this class:

$$\{a_n | \exists c_n \in D_{\sqrt{2}}, n, x_n, y_n(c_n \equiv \langle n, x_n, y_n\rangle \wedge a_n \equiv \langle n, x_n\rangle)\}$$

Analogously to the previous example, $\langle 0,1,2\rangle \in D_{\sqrt{2}}$ becomes true at level 2. At level 1, the sentence

$$\begin{aligned} \langle 0,1,2\rangle \equiv \langle 1-1,1,2\rangle \\ \wedge ((1+2^{-1})^2 > 2 \rightarrow \langle 1,1,3/2\rangle \equiv \langle 1,1,2-2^{-1}\rangle) \\ \wedge ((1+2^{-1})^2 \leq 2 \rightarrow \langle 1,1,3/2\rangle \equiv \langle 1,1+2^{-1},2\rangle) \end{aligned}$$

is also made true (the first antecedent and consequent being true and the last antecedent and consequent being false). So $D'_{\sqrt{2}}(\sigma/D_{\sqrt{2}})(c_n/\langle 1,1,3/2\rangle)$ is made true at level 2, so that at level 3, $\langle 1,1,3/2\rangle \in D_{\sqrt{2}}$ can be made true. At level 4, $\langle 2,5/4,3/2\rangle \in D_{\sqrt{2}}$ is made true, as is $\langle 3,11/8,3/2\rangle \in D_{\sqrt{2}}$ at level 5. At every finite level thereafter one more element is added to the sequence.

For all rational numerals X_0 and Y_0, the sentence $\langle 0, X_0, Y_0\rangle \in D_{\sqrt{2}}$ is made false at level 2 whenever X_0 does not designate 1 or Y_0 does not designate 2 (as both the first disjunct and the two consequents are false while one of the antecedents is true). It follows that at level 2, all triples with 0 as the first element, except from $\langle 0,1,2\rangle$, will make the conjunct $c_{n-1} \equiv \langle n-1, x_{n-1}, y_{n-1}\rangle$ false whenever c_{n-1} satisfies the condition $c_{n-1} \in D_{\sqrt{2}}$, so any choice of a triple for c_n which has 1 as the first element but is not $\langle 1,1,3/2\rangle$ will make one of the two last conjuncts false. Ergo, for all rational numerals X_1 and Y_1, the sentence $\langle 1, X_1, Y_1\rangle \in D_{\sqrt{2}}$ is made false at level 3 whenever X_1 is different from 1 or Y_1 is different from 3/2. Generalizing this reasoning, it is seen that only one triple for each possible choice of first element “gets into” the sequence.

Each time a triple $\langle N, X_n, Y_n\rangle$ makes $\langle N, X_n, Y_n\rangle \in D_{\sqrt{2}}$ true at some level, the defining formula of $\sqrt{2}$ is made true at the same level, so at the next level $\langle N, X_n\rangle \in \sqrt{2}$ is made true. Notice that as 0 is not a natural numeral and hence not in the range of the natural variables, this does not include $\langle 0,1\rangle \in \sqrt{2}$. And *vice versa*: Each time a triple $\langle N, X_n, Y_n\rangle$ makes $\langle N, X_n, Y_n\rangle \in D_{\sqrt{2}}$ false at some level, $\langle N, X_n\rangle \in \sqrt{2}$ is made false at the next level.

So as in the example with the Fibonacci sequence, the sequence is complete at level ω. That it becomes complete can be expressed in the object language with the sentence $\sqrt{2} \in \mathbb{R}$. That this sentence is true is seen as follows. It follows from the above that the sentence $\forall n_1 \exists! q_1(\langle n_1, q_1\rangle \in \sqrt{2})$ is made true at level $\omega+1$. Let N_1 and N_2 be natural numerals and Q_1 and Q_2 be rational numerals and assume that $\langle N_1, Q_1\rangle \in \sqrt{2}$ and $\langle N_2, Q_2\rangle \in \sqrt{2}$ are true. The difference between a rational number in the sequence and the next is either 0 or 2 raised to minus the index of the latter. Therefore, the following – where symbolism is as an exception used in the meta-language – holds, when we assume, without loss of generality, that $I(N_1)$ is smaller than or equal to $I(N_2)$:

$$\begin{aligned} |I(Q_1) - I(Q_2)| \leq \sum_{i=I(N_1)+1}^{I(N_2)} 2^{-i} &= 2^{-I(N_1)} - 2^{-I(N_2)} \\ &\leq 2^{-I(N_1)} \leq I(N_1)^{-1} \leq I(N_1)^{-1} + I(N_2)^{-1} \end{aligned}$$

It follows that the object language sentence $|Q_1 - Q_2| \leq N_1^{-1} + N_2^{-1}$ is true. Therefore, at level $\omega + 1$, $\sqrt{2}$ satisfies the defining formula of $\mathbb{R}$, so at level $\omega + 2$, $\sqrt{2} \in \mathbb{R}$ becomes true.

As a last expository example we will consider the sentence $\sqrt{2} <_{\mathbb{R}} \sqrt{3}$, where of course $\sqrt{3}$ is defined like $\sqrt{2}$ (the only change that is needed is to replace "2" with "3" on the right-hand side of the two inequalities). The formula $\sqrt{2} <_{\mathbb{R}} \sqrt{3}$ is short for $\sqrt{3} -_{\mathbb{R}} \sqrt{2} \in \mathbb{R}^+$ which is again short for

$$\sqrt{3} -_{\mathbb{R}} \sqrt{2} \in \{a | a \in \mathbb{R} \land \exists n, q(\langle n, q\rangle \in a \land q > n^{-1})\}.$$

This sentence is made true if both $\sqrt{3} -_{\mathbb{R}} \sqrt{2} \in \mathbb{R}$ and $\exists n, q(\langle n, q\rangle \in \sqrt{3} -_{\mathbb{R}} \sqrt{2} \land q > n^{-1})$ are true. The truth of the former follows directly from Lemma 5.6. The latter is true if there are N and Q such that $I(Q)$ is larger than the reciprocal of $I(N)$ and the sentence $\langle N, Q\rangle \in \sqrt{3} -_{\mathbb{R}} \sqrt{2}$ is true. That job is accomplished by 4 for N and 81/256 for Q. For $\sqrt{3} -_{\mathbb{R}} \sqrt{2}$ is the set

$$\{c_n | \exists n, m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in \sqrt{3} \land \langle m, y_{2n}\rangle \in \sqrt{2} \land m \equiv 2 \cdot n \land c_n \equiv \langle n, y_{2n} - x_{2n}\rangle)\},$$

and a little calculation shows that $\langle 8, 187/256\rangle \in \sqrt{3}$ and $\langle 8, 106/256\rangle \in \sqrt{2}$ are true, as is of course $8 \equiv 2 \cdot 4$ and $\langle 4, 81/256\rangle \equiv \langle 4, 187/256 - 106/256\rangle$.

5.7 Expressive weakness

There are two points that I want to make about this theory of real numbers, one negative and one positive. The negative, which is the subject of this short section, is that the theory shares the problems of expressive weakness with Kripke's theory of truth. The positive, which will be made in the following sections, is that with this theory we can, in spite of its shortcomings, make sense of Cantor's diagonal proof in a way that does not commit us to higher infinities.

Using Russell's Class, we can design a pathological almost-real-number:

$$\mathcal{O} := \{a | (\mathcal{R} \in \mathcal{R} \wedge a \equiv \langle 1, 0 \rangle) \vee (\mathcal{R} \notin \mathcal{R} \wedge a \equiv \langle 1, 0 \rangle) \vee \exists n > 1 (a \equiv \langle n, 0 \rangle)\}$$

By the third disjunct, every sentence $\langle N, 0 \rangle \in \mathcal{O}$ for a natural numeral N larger than 1 is true. In addition, every sentence of the form $\langle N, Q \rangle \in \mathcal{O}$, where N is any natural numeral and Q is a rational numeral not denoting zero, is false. However, the first conjunct of the first disjunct and the first conjunct of the second disjunct are undefined, and as a consequence, the sentence $\langle 1, 0 \rangle \in \mathcal{O}$ is undefined. We therefore have a class that is almost the real number 0, but not quite, and the sentence $\mathcal{O} \in \mathbb{R}$ is undefined.

This brings havoc to the theory. The trivial fact that the class of non-negative real numbers is a subclass of the class of real numbers cannot be expressed. That is to say that the sentence $\mathbb{R}^{0+} \subseteq \mathbb{R}$, that is

$$\forall c (c \in \mathbb{R}^{0+} \rightarrow c \in \mathbb{R}),$$

is undefined because of $\mathcal{O}$ and classes like it ($\mathcal{O} \in \mathbb{R}^{0+}$ is undefined and $\mathcal{O} \in \mathbb{R}$ is undefined, so by Strong Kleene, $\mathcal{O} \in \mathbb{R}^{0+} \rightarrow \mathcal{O} \in \mathbb{R}$ is undefined). But intuitively $\mathcal{O}$ is not a counter-example to the statement, for it is not the case that $\mathcal{O} \in \mathbb{R}^{0+}$ is true while $\mathcal{O} \in \mathbb{R}$ is not. Nor is any other class: any *possible* element of $\mathbb{R}^{0+}$ must also (potentially) be an element of $\mathbb{R}$. So the intuitive verdict is that the sentence is true. And it will be demonstrated that this intuitive verdict is not in conflict with the idea of *grounding* that guides Kripke.

Another example is that what we have just proved as Lemma 5.6 is not true in its object language form, which is

$$\forall a, b (a \in \mathbb{R} \wedge b \in \mathbb{R} \rightarrow a -_{\mathbb{R}} b \in \mathbb{R}).$$

One instance of the doubly universally quantified sentence is

$$\mathcal{O}' \in \mathbb{R} \wedge \mathcal{O}' \in \mathbb{R} \rightarrow \mathcal{O}' -_{\mathbb{R}} \mathcal{O}' \in \mathbb{R},$$

where $\mathcal{O}'$ is as $\mathcal{O}$ except that it is the membership of $\langle 2, 0 \rangle$ rather than $\langle 1, 0 \rangle$ that is made to be undefined. The class $\mathcal{O}' -_{\mathbb{R}} \mathcal{O}'$ is like $\mathcal{O}$ in that any sentence of the form $\langle N, Q \rangle \in \mathcal{O}' -_{\mathbb{R}} \mathcal{O}'$ has the same truth value as $\langle N, Q \rangle \in \mathcal{O}$. Therefore, both the antecedent and the consequent of the mentioned instance are undefined.

Examples of this sort are ubiquitous, but these two should suffice to give the general idea. Almost all universal generalizations over real numbers or classes of real numbers that we care about in analysis will come out undefined. The damaging consequences of failure of bivalence are worse than in intuitionism. The last two chapters of this dissertation are devoted to solving this problem by arguing that more sentences are to be admitted as "grounded" than Kripke does and developing precise theories that do that.

5.8 Bishop's diagonal proof

First a more immediate success story, though. Just as Cantor, Bishop (1967, 14) calls a set "countably infinite" if there is a one-one correspondence between that set and the set of integers, and just as Cantor, he then goes on to prove that the set of real numbers is not countable. But as will be shown, we can do the same and yet conclude that this does not mean that there are *more* real numbers than there are natural numbers.

Bishop's formulation of the theorem of the uncountability of the reals and its proof go as follows (quoted in full from page 25):

> **Theorem 1** Let $\{a_n\}$ be a sequence of real numbers. Let x_0 and y_0 be real numbers, $x_0 < y_0$. Then there exists a real number x with
>
> (2.22) $$x_0 \le x \le y_0$$
>
> and
>
> (2.23) $$x \ne a_n \qquad (n \in \mathbb{Z}^+)$$
>
> **Proof** We construct by induction sequences $\{x_n\}$ and $\{y_n\}$ of rational numbers such that
>
> (2.24) (i) $x_0 \le x_n \le x_m < y_m \le y_n \le y_0 \qquad (m \ge n \ge 1)$
> (ii) $x_n > a_n$ or $y_n < a_n \qquad (n \ge 1)$
> (iii) $y_n - x_n < n^{-1} \qquad (n \ge 1)$
>
> Assume that $n \ge 1$ and that $x_0, \dots, x_{n-1}, y_0, \dots, y_{n-1}$ have been constructed. Either $a_n > x_{n-1}$ or $a_n < y_{n-1}$. In case $a_n > x_{n-1}$, let x_n be any rational number with $x_{n-1} < x_n < \min\{a_n, y_{n-1}\}$, and let y_n be any rational number with $x_n < y_n < \min\{a_n, y_{n-1}, x_n + n^{-1}\}$. Then the relevant inequalities are satisfied. In case $a_n < y_{n-1}$, let y_n be any rational number with $\max\{a_n, x_{n-1}\} < y_n < y_{n-1}$, and x_n any rational number with $\max\{a_n, x_{n-1}, y_n - n^{-1}\} < x_n < y_n$. Again, the relevant inequalities are satisfied. This completes the induction.

> From (i) and (iii) follows that
>
> $$|x_m - x_n| = x_m - x_n < y_n - x_n < n^{-1} \qquad (m \geq n)$$
>
> Similarly $|y_m - y_n| < n^{-1}$ for $m \geq n$. Therefore $x \equiv \{x_n\}$ and $y \equiv \{y_n\}$ are real numbers. By (iii), they are equal.[22] By (i), $x_n \leq x$ and $y_n \geq y$ for all n. If $a_n < x_n$ then $a_n < x$, so $a_n \neq x$. If $a_n > y_n$ then $a_n > y = x$, so $a_n \neq x$. Thus x satisfies (2.22) and (2.23).

Immediately after the proof, Bishop writes that

> Theorem 1 is the famous theorem of Cantor, that the real numbers are uncountable. The proof is essentially Cantor's "diagonal" proof.

inviting the interpretation that the theorem means exactly the same thing as it does in Cantor's theory (Cantor 1891), namely that there are more real numbers than there are natural numbers.[23] We have to look in the chapter's endnotes (p. 60) to find this hint that there is something else going on:

> There is a paradox growing out of Theorem 1 which the reader should resolve. Since every regular sequence of rational numbers [i.e. real number] can presumably be described by a phrase in the English language, and since the phrases in the English language can be sequentially ordered, the regular sequences of rational numbers can be sequentially ordered, in contradiction to Theorem 1.

The solution (in our theory; to repeat, I am not making exegetical claims) is that there is an order of dependency between definitions of real numbers, and that every defined sequence of real numbers facilitates, through diagonalization, a definition of a new real number not in the sequence. Consider these examples: Let $\prec$ be the well-ordering of the rational numbers such that $x \prec y$ if x_1/x_2 and y_1/y_2 are the representations of x and y respectively as irreducible fractions with positive denominator and either $x_1 + x_2$ is less than $y_1 + y_2$ or $x_1 + x_2$ equals $y_1 + y_2$ and x_1 is less than y_1. Then define, for each natural number n, the real a_n to be the constant sequence having the n'th element of $\prec$ as every term. Next, define a to be the sequence having a_n as the n'th term, and let $x_{a,0,1}$ be the[24] diagonalization of a between 0 and 1. And then define a' to be the sequence whose first term is $x_{a,0,1}$ and where a_{n-1} is the n'th term for all n's larger than 1. Finally, let $x_{a',0,1}$ be the diagonalization of a' between 0 and 1.

[22] This means that $|x_n - y_n| \leq 2n^{-1}$ for all natural numbers n (p. 15). However, actually the equality is of no importance for the proof and is therefore ignored below.
[23] Bishop's proof is closer to Cantor's first proof in (Cantor 1874) than to the (Cantor 1891)-proof, which is more elegant and better known today.
[24] About the definite form, see below.

What we just did was to define the reals a_n first, then the real $x_{a,0,1}$, and lastly define the real $x_{a',0,1}$. A real is a rule, some rules depend on or refer to other rules, and a real defined by diagonalization depends on the reals in the diagonalized sequence. That gives us the alternative interpretation of Bishop's Cantorian theorem: however many reals have been defined, another real number, depending on them, can be defined. The class of reals is indefinitely extendable, not uncountable in the classical sense.

In the next section we will verify this by modeling Bishop's theorem and proof in our formal system. In that formal system, where we, by virtue of the classical background theory, pretend that all the classes (qua linguistic objects) exist "from the beginning", the dependency takes the form that the sentence saying that $x_{a',0,1}$ is a real number becomes true at a higher level than where the sentence asserting that $x_{a,0,1}$ is a real number becomes true, which is again at a higher level than those where the sentences for all the a_n's become true.

5.9 Diagonalization

We want to make a formal definition of "the diagonalization of a between x_0 and y_0" like we have made definitions of the Fibonacci sequence and $\sqrt{2}$. Before that can be done, there is a technicality about Bishop's proof that we have to consider. Strictly speaking there is a mistake in the proof. Bishop has to show that x is a real number, which according to Bishop's definition is a certain kind of sequence, and a sequence is again defined as a rule. But x is not given by a rule but by an infinity of choices ("let x_n be any rational number with [...]"). So it is only a real by Brouwer's definition where free choice sequences are allowed. But the proof is easily corrected, as the choices can be replaced with a rule. The basic idea I have used is to take the average of the lower and the upper bound that Bishop mentions. But this has to be qualified, for a_n may not be rational, in which case the average may not be either. But a_n is itself a sequence of rationals, so one of these can be used if only we use a element sufficiently late in the sequence: the existence of a element $q_{n,m}$ in a_n such that $q_{n,m} - x_{n-1} > m^{-1}$, where m is the position in a_n, is equivalent to $a_n > x_{n-1}$ by Bishop's definition of the standard order relation on the reals. Similarly, $a_n < y_{n-1}$ is equivalent to the existence of an $q_{n,m}$ which satisfies $y_{n-1} - q_{n,m} > m^{-1}$. So letting m be the least natural number such that $q_{n,m}$ satisfies either of those conditions, $q_{n,m}$ can play the role that a_n does in Bishop's construction. If $q_{n,m}$ satisfies the first condition, we can define

$$x_n = \frac{x_{n-1} + \min\{q_{n,m} - m^{-1}, y_{n-1}\}}{2}$$

and
$$y_n = \frac{x_n + \min\{q_{n,m} - m^{-1}, y_{n-1}, x_n + n^{-1}\}}{2}.$$
If $q_{n,m}$ only satisfies the second condition, we instead define
$$y_n = \frac{\max\{q_{n,m} + m^{-1}, x_{n-1}\} + y_{n-1}}{2}$$
and
$$x_n = \frac{\max\{q_{n,m} + m^{-1}, x_{n-1}, y_n - n^{-1}\} + y_n}{2}.$$
For this to work when $n = 1$, x_0 and y_0 have to be rational. Bishop allows them to be any real numbers. Achieving that generality would make things significantly more complicated in the present formal setting, so since it is of no essence, we will forgo it.

We therefore make the formal definition as follows: the diagonalization of A between X_0 and Y_0, written $\mathcal{D}_{A,X_0,Y_0}$, is $\{c_n | \exists\sigma(P(\sigma) \wedge \mathcal{D}'_{A,X_0,Y_0})\}$ where $\mathcal{D}'_{A,X_0,Y_0}$ is

$c_n \equiv \langle 0, X_0, Y_0\rangle$

$\vee\exists c_{n-1} \in \sigma, n, x_{n-1}, y_{n-1}, x_n, y_n, m, a_n, q_{n,m}$

$(c_{n-1} \equiv \langle n-1, x_{n-1}, y_{n-1}\rangle$

$\wedge c_n \equiv \langle n, x_n, y_n\rangle$

$\wedge\langle n, a_n\rangle \in a$

$\wedge\langle m, q_{n,m}\rangle \in a_n$

$\wedge(q_{n,m} - x_{n-1} > m^{-1} \vee y_{n-1} - q_{n,m} > m^{-1})$

$\wedge\forall m'\exists q_{n,m'}((\langle m', q_{n,m'}\rangle \in a_n \wedge (q_{n,m'} - x_{n-1} > m'^{-1} \vee y_{n-1} - q_{n,m'} > m'^{-1}))$

$\to m' \geq m)$

$\wedge(q_{n,m} - x_{n-1} > m^{-1}$

$\to (x_n \equiv \frac{x_{n-1}+\min\{q_{n,m}-m^{-1}, y_{n-1}\}}{2} \wedge y_n \equiv \frac{x_n+\min\{q_{n,m}-m^{-1}, y_{n-1}, x_n+n^{-1}\}}{2}))$

$\wedge(q_{n,m} - x_{n-1} \leq m^{-1}$

$\to (y_n \equiv \frac{\max\{q_{n,m}+m^{-1}, x_{n-1}\}+y_{n-1}}{2} \wedge x_n \equiv \frac{\max\{q_{n,m}+m^{-1}, x_{n-1}, y_n-n^{-1}\}+y_n}{2})))$

and where P is a predicate whose interpretation is the singleton of $\mathcal{D}_{A,X_0,Y_0}$. The real number X_{A,X_0,Y_0}[25] is defined from $\mathcal{D}_{A,X_0,Y_0}$ just like $\sqrt{2}$ was defined from $D_{\sqrt{2}}$, namely as
$$\{b_n | \exists c_n \in \mathcal{D}_{A,X_0,Y_0}, n, x_n, y_n(c_n \equiv \langle n, x_n, y_n\rangle \wedge b_n \equiv \langle n, x_n\rangle)\}.$$

[25] This is an exemption from the notational conventions given in Section 5.3 in two ways. First, X_{A,X_0,Y_0}, corresponding to Bishop's "x", is a class not a rational number. Second, the "A" of X_{A,X_0,Y_0} is a meta-language variable for a sequence of real numbers, and "X_0" and "Y_0" are rational numerals. The same holds for the subscripts of "$\mathcal{D}_{A,X_0,Y_0}$" and "$\mathcal{D}'_{A,X_0,Y_0}$".

The following theorem is the formal analogue of Bishop's diagonalization theorem. One final definition is needed to formulate it, namely a definition that "translates" a rational number into the real number that is "equal" to it: for a given rational numeral Q define $Q^{\mathbb{R}}$ to be the real number $\{c|\exists n(c \equiv \langle n, Q\rangle)\}$. The basic idea of the proof of the theorem is the same as in Bishop's proof, but the fact that it is about a formal language adds a significant amount of complexity, as do the corrections that were needed to make X_{A,X_0,Y_0} into a proper rule.

Theorem 5.7. *For all classes A and rational numerals X_0 and Y_0 such that $A \in \mathcal{S}_{\mathbb{R}}$ and $X_0 < Y_0$ are true sentences, also the sentences $X_{A,X_0,Y_0} \in \mathbb{R}$, $X_0^{\mathbb{R}} \leq_{\mathbb{R}} X_{A,X_0,Y_0} \leq_{\mathbb{R}} Y_0^{\mathbb{R}}$ and $\forall n, a_n(\langle n, a_n\rangle \in A \to X_{A,X_0,Y_0} \neq_{\mathbb{R}} a_n)$ are true.*

Proof. This proof is divided into four parts. In parts 2, 3 and 4, the three parts of the consequent of the Theorem are deduced. But first, in part

1. it is shown by induction that for each non-negative integer n, there are unique rational numerals X_n and Y_n such that $\langle N, X_n, Y_n\rangle \in \mathcal{D}_{A,X_0,Y_0}$, where N is the rational numeral such that $I(N)$ is n, is true and that for these constants the following holds:[26]

(a) $I(X_n)$ is smaller than $I(Y_n)$.

(b) $I(X_n)$ is larger than $I(X_m)$ for all natural numbers m smaller than n.

(c) $I(Y_n)$ is smaller than $I(Y_m)$ for all natural numbers m smaller than n.

(d) When n is positive, $I(Y_n)$ minus $I(X_n)$ is smaller than the reciprocal of $I(N)$.

Base case: Existence is trivial, as the sentence

$$\mathcal{D}'_{A,X_0,Y_0}(\sigma/\mathcal{D}_{A,X_0,Y_0})(c_n/\langle 0, X_0, Y_0\rangle)$$

is made true at level 1 and the sentence $\langle 0, X_0, Y_0\rangle \in \mathcal{D}_{A,X_0,Y_0}$ therefore at level 2.

That (a) holds follows directly from the Theorem's assumptions. Both (b), (c) and (d) are vacuously satisfied.

Uniqueness is seen as follows: For all rational numerals X_{-1} and Y_{-1}, the sentence $\langle -1, X_{-1}, Y_{-1}\rangle \in \mathcal{D}_{A,X_0,Y_0}$ is made false at level 2, as

$$\mathcal{D}'_{A,X_0,Y_0}(\sigma/\mathcal{D}_{A,X_0,Y_0})(c_n/\langle -1, X_{-1}, Y_{-1}\rangle)$$

[26] (a), (b) and (c) correspond to (i) in Bishop's proof and (d) correspond to (iii). Statement (ii) is left out and instead blended into part 4.

is made false at level 1. It follows that $\langle 0, X'_0, Y'_0 \rangle \in \mathcal{D}_{A,X_0,Y_0}$, where also X'_0 and Y'_0 are rational constants, is made false at level 3 when either X'_0 is different from X_0 or Y'_0 is different from Y_0.

Induction step: Let a positive integer n be given, and let N be the natural numeral such that $I(N)$ is n. By the induction hypothesis, there are unique rational numerals X_{n-1} and Y_{n-1} such that $\langle N-1, X_{n-1}, Y_{n-1} \rangle \in \mathcal{D}_{A,X_0,Y_0}$ is true. For these numerals it also holds that $I(X_{n-1})$ is smaller than $I(Y_{n-1})$. From the assumption that $A \in \mathcal{S}_{\mathbb{R}}$ is true, it follows that there is a unique class A_n such that $\langle N, A_n \rangle \in A$ is true.[27] It also follows that $A_n \in \mathbb{R}$ is true, from which it can be inferred that $\forall n_1 \exists! q_1 (\langle n_1, q_1 \rangle \in A_n$ is true. Therefore, for each natural numeral M there is a unique rational numeral $Q_{n,m}$ such that $\langle M, Q_{n,m} \rangle \in A_n$ is true. As $I(X_{n-1})$ is smaller than $I(Y_{n-1})$, there is some such M and corresponding $Q_{n,m}$ such that $I(Q_{n,m})$ minus $I(X_{n-1})$ is larger than the reciprocal of $I(M)$ or $I(Y_{n-1})$ minus $I(Q_{n,m})$ is larger than the reciprocal of $I(M)$. Let henceforth M designate the rational numeral which satisfies this requirement and for which $I(M)$ is smallest, and let $Q_{n,m}$ designate the unique rational numeral such that $\langle M, Q_{n,m} \rangle \in A_n$ is true.

Assume first that $I(Q_{n,m})$ minus $I(X_{n-1})$ is larger than the reciprocal of $I(M)$. In that case define X_n to be the rational numeral such that $I(X_n)$ is the average of $I(X_{n-1})$ and the minimum of $I(Q_{n,m} - M^{-1})$ and $I(Y_{n-1})$, and Y_n to be the rational numeral such that $I(Y_n)$ is the average of $I(X_n)$ and the minimum of $I(Q_{n,m} - M^{-1})$, $I(Y_{n-1})$ and $I(X_n + N^{-1})$. Then this sentence is true:

$\langle N-1, X_{n-1}, Y_{n-1} \rangle \equiv \langle N-1, X_{n-1}, Y_{n-1} \rangle$

$\wedge \langle N, X_n, Y_n \rangle \equiv \langle N, X_n, Y_n \rangle$

$\wedge \langle N, A_n \rangle \in A$

$\wedge \langle M, Q_{n,m} \rangle \in A_n$

$\wedge (Q_{n,m} - X_{n-1} > M^{-1} \vee Y_{n-1} - Q_{n,m} > M^{-1})$

$\wedge \forall m' \exists q_{n,m'} ((\langle m', q_{n,m'} \rangle \in A_n \wedge (q_{n,m'} - X_{n-1} > m'^{-1} \vee Y_{n-1} - q_{n,m'} > m'^{-1}))$

$\quad \rightarrow m' \geq M)$

$\wedge (Q_{n,m} - X_{n-1} > M^{-1}$

$\quad \rightarrow (X_n \equiv \frac{X_{n-1} + \min\{Q_{n,m} - M^{-1}, Y_{n-1}\}}{2}$

[27] In the Fibonacci and square root examples, an ordered pair/triple came into the sequence immediately after the previous ordered pair/triple was included. In this case the nth triple has to wait for two things before it can enter into the sequence, namely that triple number $n-1$ is included and that this sentence, $\langle N, A_n \rangle \in A$, is made true, i.e. that the nth real which is to be used in the diagonalization is created and included in the diagonalized sequence. Ergo, the temporal order is as described in Section 5.8.

$\quad\wedge Y_n \equiv \frac{X_n+\min\{Q_{n,m}-M^{-1},Y_{n-1},X_n+N^{-1}\}}{2}))$

$\wedge(Q_{n,m} - X_{n-1} \leq M^{-1}$

$\quad\rightarrow (Y_n \equiv \frac{\max\{Q_{n,m}+M^{-1},X_{n-1}\}+Y_{n-1}}{2}$

$\quad\wedge X_n \equiv \frac{\max\{Q_{n,m}+M^{-1},X_{n-1},Y_n-N^{-1}\}+Y_n}{2})))$

So it follows that

$$\mathcal{D}'_{A,X_0,Y_0}(\sigma/\mathcal{D}_{A,X_0,Y_0})(c_n/\langle N, X_n, Y_n\rangle)$$

is true. Hence $\langle N, X_n, Y_n\rangle \in \mathcal{D}_{A,X_0,Y_0}$ is true as well.

We need to show that (a) holds for the defined numerals, and since $I(Y_n)$ is the average of $I(X_n)$ and the minimum of $I(Q_{n,m} - M^{-1})$, $I(Y_{n-1})$ and $I(X_n + N^{-1})$, this amounts to showing that $I(X_n)$ is smaller than each of those. It follows from the assumption we are working under right now, that $I(X_{n-1})$ is smaller than $I(Q_{n,m} - M^{-1})$, and the induction hypothesis tells us that $I(X_{n-1})$ is smaller than $I(Y_{n-1})$. As $I(X_n)$ is the average of $I(X_{n-1})$ and the minimum of $I(Q_{n,m} - M^{-1})$ and $I(Y_{n-1})$, it follows that $I(X_n)$ is indeed smaller than both $I(Q_{n,m}-M^{-1})$ and $I(Y_{n-1})$. And as $I(N)$ is positive, $I(X_n)$ is smaller than $I(X_n + N^{-1})$.

From this it is also seen that $I(X_n)$ is larger than $I(X_{n-1})$ and then (b) follows from (b) of the induction hypothesis. Likewise with (c). As $I(Y_n)$ is the average of $I(X_n)$ and something which is larger than this but smaller than or equal to $I(X_n + N^{-1})$, also (d) holds.

Now assume instead the opposite, namely that $I(Q_{n,m})$ minus $I(X_{n-1})$ is smaller than or equal to the reciprocal of $I(M)$. In that case define Y_n to be the rational numeral such that $I(Y_n)$ is the average of $I(Y_{n-1})$ and the maximum of $I(Q_{n,m} + M^{-1})$ and $I(X_{n-1})$, and X_n to be the rational numeral such that $I(X_n)$ is the average of $I(Y_n)$ and the maximum of $I(Q_{n,m} + M^{-1})$, $I(X_{n-1})$ and $I(Y_n - N^{-1})$. From the assumption and the definition of M it follows that $I(Y_{n-1})$ minus $I(Q_{n,m})$ is larger than the reciprocal of $I(M)$. So the deduction under this assumption goes through analogously.

Now for the uniqueness part. It has to be demonstrated that for $\langle N, X'_n, Y'_n\rangle \in \mathcal{D}_{A,X_0,Y_0}$ to be true, X'_n has to be X_n and Y'_n has to be Y_n. As $I(N)$ is positive, $\langle N, X'_n, Y'_n\rangle \equiv \langle 0, X_0, Y_0\rangle$ is false, whatever rational numerals X'_n and Y'_n are. Therefore, all of these conjuncts have to be satisfied by suitable numerals in place of x_{n-1}, y_{n-1}, m and $q_{n,m}$ and suitable classes in place of c_{n-1} and a_n:

1. $c_{n-1} \in \mathcal{D}_{A,X_0,Y_0}$
2. $c_{n-1} \equiv \langle N-1, x_{n-1}, y_{n-1}\rangle$

3. $\langle N, a_n \rangle \in A$

4. $\langle m, q_{n,m} \rangle \in a_n \wedge (q_{n,m} - x_{n-1} > m^{-1} \vee y_{n-1} - q_{n,m} > m^{-1})$
$\wedge \forall m' \exists q_{n,m'}((\langle m', q_{n,m'} \rangle \in a_n$
$\wedge (q_{n,m'} - x_{n-1} > m'^{-1} \vee y_{n-1} - q_{n,m'} > m'^{-1})) \to m' \geq m)$

5. $q_{n,m} - x_{n-1} > m^{-1}$
$\to (x_n \equiv \frac{x_{n-1} + \min\{q_{n,m} - m^{-1}, y_{n-1}\}}{2} \wedge y_n \equiv \frac{x_n + \min\{q_{n,m} - m^{-1}, y_{n-1}, x_n + n^{-1}\}}{2})$

6. $q_{n,m} - x_{n-1} \leq m^{-1}$
$\to (y_n \equiv \frac{\max\{q_{n,m} + m^{-1}, x_{n-1}\} + y_{n-1}}{2} \wedge x_n \equiv \frac{\max\{q_{n,m} + m^{-1}, x_{n-1}, y_n - n^{-1}\} + y_n}{2})$

By the induction hypothesis, there is only one choice of numerals to substitute for x_{n-1} and y_{n-1}, if c_{n-1} is to be replaced with a class that can make both 1. and 2. true, and it has already been argued that the there is only one option for a class to take the place of a_n in 3. For a substitution for m to satisfy 4., it has to be minimal among those that satisfy a certain condition, so at most one natural numeral can. And then the uniqueness of a substitute for $q_{n,m}$ follows from the first conjunct of 4. together with 3. which implies that $a_n \in \mathbb{R}$ is true. With the unique numerals to take the places of $q_{n,m}$, x_{n-1} and m, one of the antecedents of 5. and 6. is true and one is false. The conditional with the false antecedent is true, so there is only the consequent of the other conditional left to satisfy. Whether it is the antecedent of 5. or of 6., the right-hand side of the left conjunct is now "determined", so there is only one possible choice of a constant to substitute for the left-hand side. Consequently, the same holds for the right conjunct, so X_n and Y_n are uniquely determined. This completes the induction.

2. From the induction and the definition of X_{A,X_0,Y_0}, it is easily seen that for each natural number, there is a unique rational numeral X_n such that $\langle N, X_n \rangle \in X_{A,X_0,Y_0}$ is true, where, as usual, N is the natural numeral such that $I(N)$ is that natural number (and there must be some level at which they are all true). Ergo, $\forall n_1 \exists! q_1(\langle n_1, q_1 \rangle \in X_{A,X_0,Y_0}$ is true. Furthermore, for any natural numbers n and m, where we without loss of generality assume that m is larger than or equal to n, the absolute value of $I(X_m)$ minus $I(X_n)$ is equal to $I(X_m)$ minus $I(X_n)$ (by (b)), which is smaller than $I(Y_n)$ minus $I(X_n)$ (by (a) and (c)), which is again smaller than the reciprocal of $I(N)$ (by (d)), which is finally smaller than the reciprocal of $I(N)$ plus the reciprocal of $I(M)$. Ergo the sentence $|X_m - X_n| < N^{-1} + M^{-1}$ is true. Therefore,

$$\forall n_1 \exists! q_1(\langle n_1, q_1 \rangle \in X_{A,X_0,Y_0} \wedge \forall n_2 \forall q_2(\langle n_2, q_2 \rangle \in X_{A,X_0,Y_0} \to |q_2 - q_1| \leq n_2^{-1} + n_1^{-1}))$$

is true and it follows that $X_{A,X_0,Y_0} \in \mathbb{R}$ is as well.

3. The sentence $X_0^{\mathbb{R}} \leq_{\mathbb{R}} X_{A,X_0,Y_0}$ is $X_{A,X_0,Y_0} -_{\mathbb{R}} X_0^{\mathbb{R}} \in \mathbb{R}^{0+}$. To show that it is true, the truth of two other sentences have to be demonstrated. The first is $X_{A,X_0,Y_0} -_{\mathbb{R}} X_0^{\mathbb{R}} \in \mathbb{R}$, but this follows directly from Lemma 5.6 together with the (trivial) fact that $X_0^{\mathbb{R}} \in \mathbb{R}$ is true and the demonstration in part 2 that $X_{A,X_0,Y_0} \in \mathbb{R}$ is as well. The second is

$$\forall n, q(\langle n, q\rangle \in X_{A,X_0,Y_0} -_{\mathbb{R}} X_0^{\mathbb{R}} \to q \geq -n^{-1}).$$

Since both $X_0^{\mathbb{R}} \in \mathbb{R}$ and $X_{A,X_0,Y_0} \in \mathbb{R}$ are true, all sentences of the form $\langle M, Q\rangle \in X_{A,X_0,Y_0} -_{\mathbb{R}} X_0^{\mathbb{R}}$ are true or false, so there are no undefined antecedents to worry about. Assume that it is true. Let X_{2m} be the rational numeral and L the natural numeral such that the sentences $L \equiv 2 \cdot M$ and $\langle L, X_{2m}\rangle \in X_{A,X_0,Y_0}$ are true. As $X_{A,X_0,Y_0} -_{\mathbb{R}} X_0^{\mathbb{R}}$ is the class

$$\{c_m | \exists m, l, x_{2m}, x_0(\langle l, x_{2m}\rangle \in X_{A,X_0,Y_0} \\ \wedge \langle l, x_0\rangle \in X_0^{\mathbb{R}} \wedge l \equiv 2 \cdot m \wedge c_m \equiv \langle m, x_{2m} - x_0\rangle)\},$$

$I(Q)$ is identical to $I(X_{2m} - X_0)$. It follows from (b) that $I(X_{2m} - X_0)$ is positive and hence $Q \geq -M^{-1}$ is true. From this, the desired conclusion follows.

The truth of $X_{A,X_0,Y_0} \leq_{\mathbb{R}} Y_0^{\mathbb{R}}$ is of course proved analogously. With this, the present part of the proof can be concluded: the sentence $X_0^{\mathbb{R}} \leq_{\mathbb{R}} X_{A,X_0,Y_0} \leq_{\mathbb{R}} Y_0^{\mathbb{R}}$ is true.

4. Let N be a natural numeral and A_n be a class. It follows from the assumption that $A \in \mathcal{S}_{\mathbb{R}}$ is true, that $\langle N, A_n\rangle \in A$ is either true or false, so again there are no possible undefined antecedents to worry about when trying to prove a conditional; we just need to deduce the truth of $X_{A,X_0,Y_0} \neq_{\mathbb{R}} A_n$ from the assumption of the truth of $\langle N, A_n\rangle \in A$ to be able to conclude that

$$\forall n, a_n(\langle n, a_n\rangle \in A \to X_{A,X_0,Y_0} \neq_{\mathbb{R}} a_n)$$

is true.

With the numerals meaning what they did in part 1, we have either $Q_{n,m} - X_{n-1} > M^{-1}$ true or $Y_{n-1} - Q_{n,m} > M^{-1}$ true. Assume the former (the proof is analogous in the latter case). $I(X_n)$ is smaller than $I(Q_{n,m} - M^{-1})$ and hence $I(Y_n)$, being the average of $I(X_n)$ and the minimum of a set which includes $I(Q_{n,m} - M^{-1})$, is as well. Therefore, $I(\frac{Q_{n,m}-M^{-1}-Y_n}{2})$ is positive, and we can let L be a natural numeral which satisfies the conditions that $I(L^{-1})$ is smaller than this fraction and $I(2 \cdot L)$ is larger than or equal to $I(M)$.

Let $Q_{n,2l}$ and X_{2l} be the rational numerals such that $\langle 2 \cdot L, Q_{n,2l}\rangle \in A_n$ and $\langle 2 \cdot L, X_{2l}\rangle \in X_{A,X_0,Y_0}$ are true. As $A_n \in \mathbb{R}$ is true, the difference between $Q_{n,2l}$ and $Q_{n,m}$ is at most $I(M^{-1} + (2 \cdot L)^{-1})$. And from (a), (b) and (c) we know that $I(X_{2l})$ is smaller than $I(Y_n)$. Therefore, $I(Q_{n,2l})$ minus $I(X_{2l})$ is larger than $I(Q_{n,m})$ minus $I(M^{-1})$ minus $I((2 \cdot L)^{-1})$ minus $I(Y_n)$ which is again larger than $I((2 \cdot L)^{-1})$. Ergo the sentence $Q_{n,2l} - X_{2l} > (2 \cdot L)^{-1}$ is true. Hence $\exists n, q(\langle n, q \rangle \in X_{A,X_0,Y_0} - A_n \wedge q > n^{-1})$ is true. From Lemma 5.6 it follows that also $X_{A,X_0,Y_0} - A_n \in \mathbb{R}$ is true, so $X_{A,X_0,Y_0} - A_n \in \mathbb{R}^+$ and hence $X_{A,X_0,Y_0} >_{\mathbb{R}} A_n$ and therefore $X_{A,X_0,Y_0} \neq A_n$ are as well. □

The content of the Theorem is that the class of real numbers is uncountable in Bishop's sense of that word. He has defined it in the same way as Cantor, namely as the absence of a one-one correspondence with the natural numbers. But for Bishop, and for us, a one-one correspondence is, like any other function, a rule. So what the Theorem really says is just that it is impossible to describe a sequence that contains all the reals and only reals.[28] Any sequence that purports to do so will through diagonalization give rise to a new real number which is not included in that sequence. So the class of real numbers is indefinitely extensible. At no point in time can all the reals have been defined.

That is what can be seen from the object language. But the object language is described in a classical meta-language, and that fact affords us an alternative point of view. For it is easy to see that the set of classes S such that $S \in \mathbb{R}$ is a true sentence, is classically countable. This follows simply from the fact that the vocabulary of the language is countable, which implies that there are only countably many finite combinations of the elements of that vocabulary, and the classes constitute a subset of the set of those combinations. The difference between Cantor's and Bishop's concepts of uncountability is made apparent by the fact that Bishop's reals are deemed uncountable by the object language but countable by the meta-language. The class of real numbers is indefinitely extensible but forever countable.

We should try to get a little more clear on what it means that the reals are "indefinitely extensible", as that phrase has been defined in different ways and there is disagreement about what kind of phenomenon it refers to. Russell's (1907, 36) definition of what he calls "self-reproductive" is often taken to be equivalent to Dummett's (1993, 441) definition of "indefinitely extensible", so it interesting that we here have something that satisfies the latter definition but does not satisfy (the precise wording of) the former. This is what Russell writes:

> [T]here are what we may call *self-reproductive* processes and classes. That is, there are some properties such that, given any

[28]Describing a sequence that contains all the reals is no problem: take a sequence that contains all the classes.

> class of terms all having such a property, we can always define a new term also having the property in question. (Original emphasis)

Given $\mathbb{R}$ we can *not* define a new term also having the property of being a real number, for the class $\mathbb{R}$ does not in itself give us a sequence of reals to which diagonalization can be applied. To get a definition that captures what we are dealing with here, we need to add the crucial word "definite", as Dummett does:

> An indefinitely extensible concept is one such that, if we can form a definite conception of a totality all of whose members fall under that concept, we can, by reference to that totality, characterize a larger totality all of whose members fall under it.

Much confusion has surrounded Dummett's use of the word "definite", for in this context it is not clear that its meaning is definite. We, however, are in a position to make it precise, namely as meaning "can be ordered in a sequence". Thus understood, the reals are indefinitely extensible: if a collection, all of whose members fall under the concept of *real number*, can be ordered in a sequence, we can by reference to that totality characterize a larger collection – the reals of the sequence plus its diagonalization – all of whose members fall under the concept of *real number*.[29]

It is interesting that the indefinitely extensible concept of real numbers can be "collected" in an intensional class which is a sub-class of a class characterized by a concept that is *not* indefinitely extensible, namely a universal class characterized by a tautologous concept.

5.10 Comparison with Skolem's paradox

That the diagonalization theorem can go through in a setting where there is not, seen from an "external point of view", uncountably many reals, is not in itself something new. That is also the point of Skolem's paradox[30], so a comparison between this and the result above is in order.

The so-called paradox consists in the fact that countable first-order axiomatizations of classical set theory have non-standard models in which the set

[29] Under this definition, the natural numbers are not indefinitely extensible, for they can be ordered in a sequence. However, a more liberal – but in my opinion still reasonable – notion of indefinite extensibility can be added according to which they are, namely be using the stricter interpretation of "definite collection" as meaning "all elements of the collection have been constructed". Then the natural numbers are indefinitely extensible with the "diagonalizer" being the operation of taking the maximal element of the collection and applying the successor function to it.

[30] The primary source is (Skolem 1922). See also (Benacerraf and Wright 1985).

of real numbers is countable, even though the formal sentence which says (or would normally be interpreted to say) that it is uncountable, is true in those models. That is possible because the formal sentence can be true merely due to the absence of a *witness* to countability. The formal sentence asserts the non-existence of a one-one correspondence between the natural numbers and the reals. Such a correspondence is itself a set, and this set – the witness to the countability – is what is absent from the models in question.

The theorem that the reals are uncountable is of course provable in classical set theory, so Skolem's paradox shows, just as well as the present result, that diagonalization does not on its own suffice to demonstrate that there are more reals than natural numbers. But there is an important difference. In the classical case it seems like nothing more than an anomaly; an indication that first-order axiomatizations are imperfect.[31] In that context it cannot be regarded as a reasonable aspect of the theory, for classical set theory is a theory about timeless sets – the theory about *all* the timeless sets that *could* exist. So if there is an external perspective from which it can be seen that there could be a one-one correspondence between the reals of the model and the natural numbers, then this correspondence ought to be in the model. Hence, if it is not, it is deemed a "non-standard" model. (I do not want to overstate this point though. I am not claiming that we here have another argument against classical mathematics on top of those made in Chapter 1, for the classical mathematician can avoid the problem by using second-order logic, or by denying that she is restricted in her expressive power to what is invariant in all models of some set of axioms.)

In our theory on the other hand, the result is just natural, for here the objects are defined in time in a never-ending process, so the fact that there is not at any point a completed correspondence between the natural numbers and the reals is well motivated by the background philosophy.

[31] That is the central point of (Shapiro 1991).

Chapter 6

Compositional supervaluation

We will now attend to the expressibility problem of too many sentences coming out undefined. It should be clear that the problem is virtually the same in Kripke's truth theory (see the end of Section 5.2) and our theory of classes (see Section 5.7). It does therefore not seem overly optimistic to hope that if we study the problem in the former case, we will afterwards be able to transfer insights and solutions to the latter case. As there is an extensive literature on theories of truth that can be drawn on and used as inspiration, and hardly any literature on classes relevant to the one presented here, this seems like a good methodology. In the present chapter we will therefore focus on what I call "Gupta's Challenge", which does not in itself have a direct bearing on the problems for class theory explained in Section 5.7, but nevertheless turns out to provide inspiration that will eventually lead us to a solution to the class theoretical problems.

This chapter is therefore about a modification of Kripke's theory. It is not the modification that I am ultimately going to endorse; it is only an intermediate step. It is a strengthening of the basic version of Kripke's theory in the sense that every sentence that is true (false) in the basic version is true (false) in the modification and more sentences are made true and false. However, not quite enough sentences are made true or false, I will argue in the end of this chapter. My dialectical strategy is to first convince you, the reader, that such a strengthening is necessary to capture everything that should be recognized as true or false, and then persuade you that the theory of this chapter is unstable in the sense that if those extra truths and falsities are accepted, even more truths and falsities have to be. I believe that we should accept a non-compositional theory, which will be described in Chapter 7. That is a controversial claim, and that is why I feel the need to first demonstrate the shortcomings of a compositional approach.

We will continue to put brackets around certain issues, namely TIC and TAP, the use of transfinite ordinals and the reliance on classical set theory on the meta-level, and return to these issues in Chapter 7.

6.1 Gupta's Challenge

The modification to be presented in this chapter is inspired by the following scenario, which Gupta (1982) presented in order to criticize Kripke's theory: assume that the sentences

A1: Two plus two is three
A2: Snow is always black
A3: Everything B says is true
A4: Ten is a prime number
A5: Something B says is not true

are all that is said by a person "A", and the sentences

B1: One plus one is two
B2: My name is B
B3: Snow is sometimes white
B4: At most one thing A says is true

are all that is said by a person "B". The sentences A1, A2, and A4 are clearly false and B1, B2, and B3 are clearly true. So it seems unobjectionable to reason as follows: A3 and A5 contradict each other, so at most one of them can be true. Hence at most one thing A says is true. But that is what B says with his last sentence, so everything B says is true. This is again what A says with A3 and rejects with A5, so A3 is true and A5 false.

But counterintuitively the basic version of Kripke's theory tells us that A3, A5, and B4 are all undefined. The reason is that the evaluation of A3 and A5 awaits the determination of B4, which in turn cannot receive a truth value before A3 or A5 do.

One way to obtain the intuitively correct truth values is to swear allegiance to one of the theories that assign truth values in a holistic manner.[1] But that should not be necessary; the truth of B4 and A3 and the falsity of A5 are intuitively *grounded*. The truth of B4 is intuitively grounded merely in the facts that A1, A2, and A4 are false and that A3 and A5 are contradictory. No

[1]See for example (Walicki 2009), according to which truth values are not assigned in a stage-by-stage process as in Kripke's and Gupta's theories, but "simultaneously". A rough formulation of the theory is that any assignment of truth values that satisfies certain compositional demands and minimizes the number of sentences to be declared undefined is acceptable. See also (Wen 2001).

specific assignment of truth values to A3 and A5 is presupposed. However, when B4 has been assigned the value of truth, the truth of A3 and the falsity of A5 are subsequently grounded in the truth of B1–B4.

Slightly more precisely: The fact *A3 is true* is grounded in the fact *B4 is true* plus some other facts. The fact *A5 is false* is grounded in the same facts. The fact *B4 is true* is in turn grounded in the fact *A3 and A5 are contradictory* plus some other facts. This last fact can be grounded in facts about the meaning of A3 and A5 alone. Hence the grounding relation does not violate the conditions of irreflexivity and asymmetry that must clearly hold for such a relation.

Or in yet other terms: it should be possible to *first* establish the fact that A3 and A5 are contradictory, *then* make B4 true, and *after* that make A3 true and A5 false.

All nine sentences can intuitively be assigned proper truth values (true or false; the value of "undefined" will henceforth be called a truth value but not a *proper* truth value) via a bottom-up process. Therefore, these sentences constitute a challenge to any theory which purports to be about grounded truth, such as Kripke's. Let us call it "Gupta's Challenge". Kripke (1975, 706) defines "grounded", as a predicate applicable to sentences, as meaning that the sentence has a proper truth value in the minimal fixed point. One thing that Gupta's Challenge teaches us is that the minimal fixed point of the strong Kleene valuation schema does not fully capture the informal notion of groundedness.

Note also that the intuitive reasoning for the truth values of the nine sentences does not employ any "dangerous" principles, i.e. principles that could lead to inconsistency if used in other situations. The only principle used here that goes beyond what is validated by the basic version of Kripke's theory is the inference rule, *from the inconsistency of two sentences, conclude that at most one of them is true*, which is perfectly benign. Hence, there does not seem to be any good excuse for a theory that delivers another result.

For the purposes of this chapter I will take two of the success criteria for a theory of truth to be that it only delivers truths and falsities that are intuitively grounded (this will be referred to as "groundedness-acceptable") and that it can handle Gupta's Challenge adequately. A third success criterion for a theory of truth, to be introduced in Section 6.3 and there argued to be *prima facie* reasonable, is that it is compositional (i.e. that the connectives and quantifiers are truth-functional). With respect to the first of these success criteria, the basic version of Kripke's theory is on the right track. But Gupta's Challenge shows that it does not go far enough.

In the next section I elaborate on the subject of grounding, concluding that some kind of supervaluation is compatible with the demand for grounding. In Sections 6.3 and 6.4 it is argued that Kripke's form of supervaluation and revision theory-style supervaluation are not. From Section 6.5 onwards, an alternative form of supervaluation is presented.

6.2 Grounding

The notion of grounding is more vague than the concepts of TIC and TAP, but it will have to do for now. Trying to be completely philosophically precise about a formal theory that will ultimately be discarded is not worth the trouble. As already stated, we will aim for more precision in Chapter 7.

A simplified explanation of what grounding should consists of in a hierarchical theory like Kripke's is the following: We begin with a basis consisting of non-semantical facts. Then we make all those sentences that correspond to those facts true. Now there are a number of semantical facts in addition to the non-semantical facts, namely, facts about which sentences are true. Then we make all those sentences that can be grounded in the enlarged set of facts true. And so on.

This idea can already be found in Tarski's theory of truth.[2] With his basic theory, Kripke contributed two things. First, he showed how to put the different levels of this iteration into one language (although by his own admission (1975, 714) the success here is only partial). Second, he showed how to capture formally the intuitive verdict that the idea of grounding provides in the case of the Watergate example and structurally similar cases.

The reason why the explanation two paragraphs back is a simplification is that the basis is not completely devoid of semantic facts. For instance, facts concerning what sentences are *about* are semantic facts, and they have to be included in the basis, if the truth values of the Watergate sentences are to come out right (see the definitions of the predicates N and J in Section 5.2). The only semantic facts that cannot be in the basis are those that are defined by the iteration. All facts available at a given level can be used for grounding.

This brings us back to Gupta's Challenge. That there are two sentences that cannot both be true is a semantic fact that is not defined by the iteration, but is available from the outset.[3] Hence, when A1, A2 and A4 are made false,

[2]See (Tarski 1933) and (Tarski 1944).

[3]Or rather: may be. It is a fact from the outset that ϕ and $\neg\phi$ cannot both be true. If ψ becomes false at, say, level 7, then at level 7 it also becomes a fact that $\psi \vee \phi$ and $\neg\phi$ cannot both be true.

the fact of those falsities joins an already existing fact about A3 and A5 being inconsistent. The necessary facts are then in place to make B4 true. That is what the basic version of Kripke's theory misses: Gupta's Challenge and structurally similar cases.

The formal iteration of Kripke's basic theory only makes use of 1) non-semantic facts, 2) the semantic facts that can be encoded in the interpretation function which provides the semantics of the constants and the ordinary predicates of the language, and 3) semantic facts consisting of the actual assignment of truth values that have been made at previous levels. One thing is left out with respect to the intuitive iteration he is trying to formalize: facts about which combinations of truth values are possible (given the other information).

This calls for a form of supervaluation: if a particular sentence is true relative to all possible assignments of truth values (given all other facts available at a given level), then that is a fact in which the truth of that sentence can be grounded. Hence, the attempt that Kripke makes to formulate a supervaluation version of his theory is well motivated. Unfortunately, that attempt goes too far to be groundedness-acceptable: in some cases it declares that a sentence is true given all possible assignments of truth values, when in fact it is not. That is one of the issues to be discussed in the next section.

For a hierarchical truth theory to be groundedness-acceptable it has to be the case that every time a sentence is made true or false at a level, there is an already established (negative) fact for it to be grounded in. There may be facts about which combinations of truth values are possible and which are impossible that are established prior to the sentences in question getting specific truth values, and making sentences true grounded in such a fact is acceptable – but of course only if it is a genuine fact! If the class of possibilities considered is too narrow, we get "false positive" verdicts about impossibilities of combinations, and using such verdicts will make a theory unacceptable.[4]

[4] An alternative way to characterize the deficiency in Kripke's method, suitable for the classical mathematician, is to consider the following framework for arriving at an interpretation of the truth predicate by iteration. Instead of adding true and false sentences at each level, we remove possible combinations of truth values. So where Kripke's levels each consists of an ordered pair of true and false sentences, the alternative is to have each level consist of a set of functions from the set of sentences to the set of the three truth values, representing those combinations that are not (yet) ruled out. Thus, the first level consists of the set of all such functions, and at each level functions are removed, based on the information encoded in the previous set of functions. Kripke's theory could be reformulated in this more general framework. But it would be with a jump rule that was only sensitive to information consisting in it being the case for some sentences that all remaining functions assign the sentence the value true or that all remaining functions assign the sentence the value false. That way, Kripke's method is quite simplistic, only utilizing a fraction of the information that a more sophisticated jump rule might make available. (I have not developed this idea further, partly because I don't see how to make it constructivistically acceptable.)

6.3 Kripke's supervaluation

In this section I discuss Kripke's supervaluation versions of his theory and in the next, Gupta's own revision theory of truth, in order to make the case that these theories are not successful in the stipulated senses of being groundedness-acceptable, handling Gupta's Challenge adequately and resulting in a compositional semantics.

The simplest version of supervaluation considered by Kripke is the one in which sentences are supervaluated as true (false) if they are evaluated as classically true (false) relative to all total extensions of the evaluations at a given level. More precisely: For each classical interpretation of the truth predicate (i.e. a subset of the domain) $\mathcal{E}$, let $T_{\mathcal{E}}$ be the set of classically true sentences relative to $\mathcal{E}$ (and a model), and let $F_{\mathcal{E}}$ be $\mathcal{S} \smallsetminus T_{\mathcal{E}}$, the set of classically false sentences relative to $\mathcal{E}$ (and that model). This simple supervaluation jump takes a partial interpretation (T, F) of the truth predicate to another partial interpretation (ST, SF) where ST (SF) is the set of sentences ϕ such that for all $\mathcal{E}$, if $T \subseteq \mathcal{E} \subseteq \mathcal{S} \smallsetminus F$ then $\phi \in T_{\mathcal{E}}$ $(\phi \in F_{\mathcal{E}})$.[5]

This version has no effect upon the sentences of Gupta's example compared to the basic version. The reason is that at every level there is, among the total evaluations quantified over when making the supervaluation, an evaluation in which both A3 and A5 are true, so B4 does not become true.

This problem can be remedied by tweaking the theory a bit. Instead of quantifying over all total extensions of the given evaluation, we can restrict attention to the maximally consistent ones: add the proposition that $\mathcal{E}$ is maximally consistent as a conjunct to the antecedent of the condition for ϕ above. Since there is no maximally consistent total evaluation in which both A3 and A5 are true, B4 becomes true at level 2 in this version of the theory; and then at the next level, A3 is made true and A5 false.

This does not really solve the problem, however. After the theory has been tweaked, Gupta can (and does) tweak his example to make the problem reappear by replacing A3 and A5 with A3* and A5*:

A3*: "Everything B says is true" is true
A5*: "Something B says is not true" is true

These sentences are not contradictory in the proper technical sense, only intuitively, so again B4 does not become true.

Taking a step back from the issue of which sentences are given which truth values, there is a more fundamental reason for rejecting the supervaluation

[5]See (Kripke 1975, 711).

versions of Kripke's theory. As argued above, the basic version of the theory is groundedness-acceptable: a sentence is declared true when the state of affairs expressed by the sentence can be grounded as true in the model or a lower level. This is not the case in the supervaluation versions, which declare a sentence true when the state of affairs expressed by the sentence is the case according to a class of fictions. Let me explain.

In the previous section it was concluded that if a sentence is true relative to all possible assignments of truth values (given all other facts available at a given level), then that is a fact in which the truth of that sentence can be grounded. However, Kripke's form of supervaluation does not check that the sentence is true relative to *all* possible assignments of truth values before the sentence is made true. For the final evaluation (the fixed point) must surely be considered a possible evaluation – what is actual is possible. And the final evaluation is, in the supervaluation versions just as in the original, non-total (in the presence of vicious self-reference). So the total evaluations quantified over in the supervaluation are not all the possible assignments, for they do not include the evaluation that ends up being the actual evaluation. The total evaluations are, rather, expressions of the fiction that we are in a bivalent setting when in fact we know that we are in a trivalent one. Therefore it is not sufficient to consider what is true in these; doing so gives "false positive" verdicts on matters of which combinations of truth values are possible. Ergo, Kripke's supervaluation theories are not groundedness-acceptable.

A concrete symptom of this philosophical problem is that Kripke's method for supervaluation declares all classically valid sentences true. This is very misleading in a semantics that is strictly weaker than classical semantics. A prominent example is that the disjunction of the Liar and its negation is true even though neither of the disjuncts is made true. When taking the step from bivalent to trivalent semantics, it seems clear that the property *from the information that a disjunction is true, it is possible to conclude that at least one of the disjuncts is as well*, is more important to preserve than the property *every disjunction of a sentence with its negation is true.* The latter can hardly be considered a desideratum at all (given that there are sentences ϕ such that neither ϕ nor $\neg\phi$ is true); it rather seems like dishonesty. This is why it *seems* reasonable to demand of a truth theory that it is compositional.

The restriction to maximally consistent evaluations does not prevent the disjunction of the Liar and its negation from being made true. Neither does any other restriction, for quantifying over fewer evaluations can only make more sentences true or false. In short, every version of Kripke-style supervaluation results in a non-compositional semantics.

Kripke's basic theory does not go far enough to capture all that is intuitively grounded, while his supervaluation versions go too far. There is a middle ground to be seized, one in which supervaluation is employed, but over trivalent evaluations.

6.4 Revision theory

We obtain the intuitively correct truth values for A1–A5 and B1–B4 if we accept a revision theory of truth. One version of such a theory[6] is that there are a number of best candidates for the extension of truth, namely the extensions that appear again and again when the revision process has settled into a loop. Define $T^\alpha(\mathcal{E})$ by recursion on the ordinal α as follows:

$$T^\alpha(\mathcal{E}) = \begin{cases} \mathcal{E} & \text{if } \alpha = 0 \\ T_{T^{\alpha-1}(\mathcal{E})} & \text{if } \alpha \text{ is a successor ordinal} \\ X^\alpha \cup ((\mathcal{S} \setminus Y^\alpha) \cap \mathcal{E}) & \text{if } \alpha \text{ is a limit ordinal} \neq 0 \end{cases}$$

where

$$X^\alpha = \left\{ \phi \;\middle|\; \exists \beta < \alpha \left(\phi \in \bigcap_{\beta \leq \gamma < \alpha} T^\gamma(\mathcal{E}) \right) \right\}$$

and

$$Y^\alpha = \left\{ \phi \;\middle|\; \exists \beta < \alpha \left(\phi \notin \bigcup_{\beta \leq \gamma < \alpha} T^\gamma(\mathcal{E}) \right) \right\}.$$

The set $\mathfrak{B}$ of best candidates for the set of truths is the set of evaluations B, such that for some interpretation $\mathcal{E}$, for all ordinals α, there is an ordinal $\beta > \alpha$ such that $B = T^\beta(\mathcal{E})$.[7]

These best candidates give the correct truth value to a lot of sentences that, according to intuition, are unproblematically true or false. So for example A1–A5 and B1–B4 (and the *-variant) have the right truth values in all of the best candidates, and every sentence that is true (false) in the basic version of Kripke's theory is also true (false) in the best candidates. Yet, arguably, none of the *best* candidates are *good*. We are right back at the problems that drove Tarski to "ban" self-reference: the Liar is true in some of the best candidates and false in others, both of which are bad, while "The Liar is true" has the opposite truth value.[8] Ergo, the T-schema is not validated, so according to

[6] The aim of this section is to see whether the *machinery of* revision theory can be used to formulate an acceptable theory of grounded truth in the sense of section 6.2. I am not engaging with revision theory as it is interpreted philosophically by Gupta, Belnap and Herzberger, for their aim is quite different.

[7] See (Gupta 1982, 44–45).

[8] At least that is the case with the "successor stage candidates". With some of the limit stage candidates, it is not. But in these, the truth values of some sentences are just the arbitrary value of the initial evaluation, so any advantage these candidates may have over the successor stage candidates is purely a result of arbitrariness.

Tarski's (1944, 344) adequacy condition, these candidate truth predicates are not truth predicates at all.

The advantages of revision theory over Kripke's theory are achieved using techniques that are not groundedness-acceptable, for the construction of each of the sequences is not based on established facts (i.e. the model) alone, but also on an initial evaluation that is not fact but fiction.

A different version of revision theory is that the "real" truths are the stable truths, and the "real" falsehoods are the stable falsehoods. A sentence ϕ is stably true if $\phi \in \bigcap \mathfrak{B}$ and stably false if $\phi \notin \bigcup \mathfrak{B}$.[9] By supervaluating over all initial evaluations and all the best candidates they result in, we obtain *one* definite set of truths and *one* definite set of falsehoods. However, this method also gives us sentences that are neither. Again, we have a trivalent semantics.

However, this supervaluation version of revision theory is also not groundedness-acceptable. The problem that the construction is not based on established facts but on fictions remains. There is an interesting catch-22 phenomenon here. The best candidates for the extension of truth are unacceptable because they are each based on just one initial evaluation which is a hypothesis and not fact. Supervaluations can be seen as an attempt to resolve this problem. The rationale would be that while *one* evaluation is a hypothesis, what is true relative to *all* possible evaluations is based upon facts; so the truths and falsehoods that are common to all possible evaluations are based on facts, no matter what happens to be factual, and thus *they* are groundedness-acceptable. However, this rationale does not hold, for the supervaluation gives rise to a trivalent semantics, and when one accepts such a semantics, one can no longer hold the set of all bivalent truth-value ascriptions to represent all possible states of affairs. Herein lies the catch-22: In the attempt to consider all possible states of affairs, one has to admit that they are not all the possible states of affairs after all. For a supervaluation theory to be acceptable, the kind of evaluations (bivalent, trivalent) quantified over must be of the same kind as the possible outcomes of this supervaluation.[10] I will propose such a theory.

In addition to this problem of philosophical justification, the supervaluation version of revision theory shares the problem of lack of compositionality with the supervaluation version of Kripke's theory. There are disjunctions that are declared true even though none of the disjuncts are, for example the one with the Liar and its negation.

[9] See (Gupta 1982, 46).

[10] I only have the present context, i.e. theories of truth, in mind with this claim. There may be other fields, for example the semantics of vagueness, where the use of supervaluations to go from a set of one kind of evaluations to an evaluation of another kind makes good sense.

On top of all this, the revision theory does actually not really meet Gupta's Challenge. For the solutions to the two versions considered up to now depend on the number of iterations of the truth predicate in the analogue of A3 being identical with the number of iterations of the truth predicate in the analogue of A5.

The reason that the theory leads to the intuitively correct result in the original version of the Challenge is as follows: No matter what the initial evaluation is, the six sentences without the truth predicate receive the right value at level 1 and maintain them thereafter. Then, no matter what the value of B4 is, one of the sentences A3 and A5 becomes true and the other false. Therefore, B4 becomes true, which then causes A3 to become true and A5 to become false.

Adding a truth predicate to both A3 and A5 does not change much. It just means that for A3* and A5* there is an extra delay of one level before the truth values settle into the right ones.

But if, for example, A5 is replaced with A5* while A3 is retained, the truth values for the sentences do not stabilize for all initial evaluations; for then the two sentences draw their values from different levels, so to speak, and may become true simultaneously, causing B4 to become false. This table shows what happens when the initial evaluation is one that assigns falsity to all the sentences involved (the pattern in levels 1–3 is repeated in levels 4–6 and ad infinitum):

Level	0	1	2	3	4	5	6	⋯
A3	$\bot$	$\bot$	$\top$	$\top$	$\bot$	$\top$	$\top$	⋯
A5*	$\bot$	$\bot$	$\top$	$\bot$	$\bot$	$\top$	$\bot$	⋯
B4	$\bot$	$\top$	$\top$	$\bot$	$\top$	$\top$	$\bot$	⋯

The intuitive argument presented in Section 6.1 is equally strong no matter how many additional truth predicates are applied to some of the sentences. Therefore this failure shows that revision theory does not get to the heart of the problem.[11]

[11] Variants of the revision theory with different limit rules are proposed in (Herzberger 1982) and (Belnap 1982). According to Gupta, a sentence that has not reached stability at a given limit ordinal should revert to its truth value in the initial evaluation. Herzberger suggests that it should revert to falsity, which will not help. Belnap thinks that we should consider all possible limit rules as long as they retain the truth values of sentences that have stabilized. He further believes that we should quantify over all such rules, and only consider a sentence to be true or false if it stabilizes as such under all rules. On this proposal, the A3/A5* version will also not arrive at the intuitively correct values, as one of the limit rules quantified over is the one starting with universal falsity and always reverting to falsity. Technically it is possible to deal with the Challenge within the framework of revision theory, namely by using "fully varied revision sequences"; see pages 168 and 228 in (Gupta and Belnap 1993). Then for each revision sequence there would be a limit ordinal where the "right" values

It has been argued above that we can safely infer that at most one of A3 and A5 can be true. To conclude from this that at most one of A3 and A5* (this example is easily generalizable) can be true, all that is needed is that A5 is true iff A5* is true. But that is merely an instance of that version of the T-schema (see Section 5.2) which holds in the supervaluation version of revision theory and in all versions of Kripke's.[12]

From the considerations of this and the previous section we can draw the conclusion that Kripke's supervaluation theories share three problems with the supervaluation version of revision theory. They both result in non-compositional semantics; they both only quantify over two-valued evaluations, but deliver three-valued evaluations; and both are inadequate to handle all versions of Gupta's Challenge. A version of Kripke's theory with a different form of supervaluation can solve the first two problems and make progress on the third.

6.5 The alternative

Before presenting this alternative form of supervaluation, I need to address the issue of the semantics of the truth predicate: for each of the three possible truth values of a sentence ϕ, what should the truth value of the sentence $T(c_\phi)$, claiming truth of ϕ, be? The naïve answer is that $T(c_\phi)$ should be true when ϕ is true, false when ϕ is false, and also false when ϕ is undefined. However, general use of that rule leads to paradox. If the Liar is undefined, the sentence claiming that the Liar is true would become false, and then the negation of that sentence, which is the Liar itself, would become true.

Kripke responds to this problem by not using the naïve rule under any circumstances. Instead he always has $T(c_\phi)$ undefined when ϕ is undefined. Use of this rule is seemingly required to get the monotonicity which is crucial to his construction. Otherwise we run the risk that at some stage $T(c_\phi)$ is made false because ϕ is (still) undefined, and then at a later stage ϕ becomes true and hence $T(c_\phi)$ becomes true. What I will show is that in some cases, the naïve rule can be safely employed. We will experiment with this in order to get the intuitively correct result for Gupta's Challenge.

---

would be assigned and thereafter kept. However, this is to search for consistent evaluations of sentences (see footnote 1), not to make a sentence true when there is a fact it can be grounded in.

[12] The A3/A5* version of the Challenge is discussed by Gupta and Belnap (1993, 228) who claim that "The intuitive argument [. . .] no longer goes through, since now A's statements [A3] and [A5*] do not contradict each other. To show that [A3] and [B4] are true one needs to appeal to an instance of the T-step, e.g., [A5* iff A5] which is not validated [...]". I beg to differ: the *intuitive* argument does go through. Gupta and Belnap provide no philosophical argument in the text, but simply reiterate the consequences of their theory for the case at hand.

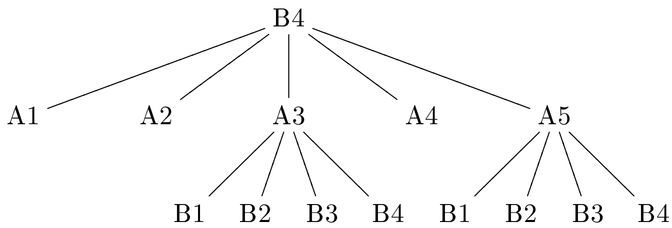


Figure 6.1

To perform supervaluation in a groundedness-acceptable way, we need to consider all possible evaluations, i.e. all possible extensions of the evaluation at some given level. Here is a way that this can be done in the case of Gupta's sentences using trees.[13] Consider the tree depicted in Figure 6.1. Here B4 is at the top (the "root") and at the nodes below it are the sentences on which the truth value of B4 depends. (I am here relying on an intuitive notion of dependency. It will be replaced by a precise definition in the next section.) Below A3 and A5 are, in turn, the sentences they depend upon. The different possible evaluations correspond to the different ways truth values can be assigned to the nodes, in such a way that nodes with other nodes below them are assigned truth values in accordance with the strong Kleene scheme based on the values of those nodes. The set of all such possible evaluations will reveal that certain combinations of truth values are impossible, *in casu* the combination of A3 being true and A5 being true. This corresponds to the crucial step in the intuitive argument for the truth values of the nine sentences: inferring that A3 and A5 contradict each other independently of what their actual truth values are.

The nodes with the sentences B1–B3 should of course be assigned the value $\top$ and the nodes with A1, A2 and A4, the value $\bot$. If we assign $\top$ to the two end nodes with B4, then under the compositional rules we should assign $\top$ to A3 and $\bot$ to A5 and therefore the root should get the value $\top$. If we instead assign $\bot$ to the end nodes with B4, the assignment to A3 and A5 should be the other way around but again the root gets a $\top$.

Lastly, we can assign the value of undefined to B4 at the bottom. Recall that A5 is "Something B says is not true" and that everything else that B says is true. So when B4 is undefined and *a fortiori* not true, A5 is intuitively true. Similarly, A3 ("Everything B says is true") is intuitively false. If we assign these values to the two nodes, the root again becomes true. So in all of the possible cases, B4 becomes true and we can supervaluate it as such.

[13]Using trees in the formulation of theories of truth has, as far as I know, only been done by Davis (1979), and he merely used it to provide an equivalent formulation of the basic version of Kripke's theory (more on this in section 6.10).

Note that the intuitive argument for the truth values of A3, A5 and B4 does *not* rely on a hidden premise that these sentences have proper truth values. Therefore this seems to be a reasonable thing to do.

However, were we to follow rules analogous to those of Kripke's basic theory, the assignment of + to the end nodes with B4 would result in also A3, A5 and the root receiving the value +. The rule in question is the weak rule for truth, that whenever a sentence ϕ is undefined, so is $T(c_\phi)$. What I just did amounts to letting $T(c_\phi)$ be false when ϕ is undefined in one of the evaluations quantified over when supervaluating. In the current setting, this can be done in some situations without the danger of making $T(c_\phi)$ false because ϕ is undefined at some stage and then having ϕ become true at a later stage. *The situations are those where a node with $T(c_\phi)$ has the root sentence below it with the value +.* (This criterion needs to be generalized, but for expository reasons that is postponed to Section 6.8.)

The reason is simple: We are only using the strong rule in an evaluation which is considered together with several other evaluations in a supervaluation. In other of those evaluations the end node with ϕ is assigned $\top$, and only when those evaluations assign the same value to the root do we go ahead and assign the root sentence an actual truth value. In this way we are, so to speak, safeguarding ourselves against the possibility that a later stage in the iteration of the jump rule will contradict what we based our truth-value assignments on at the present stage. That is why this limited use of the strong truth rule does not lead to inconsistency.

The use of the rule is restricted in two ways that should be distinguished. First, it is only used in the evaluations that are quantified over. That is, if some sentence ϕ is left undefined at a given level, because the supervaluation for it did not deliver a definite result, that is not taken as a sufficient basis for making $T(c_\phi)$ false. Second, the strong rule is limited in its application to nodes in a tree that has the root sentence appearing again below it, where it is assigned +. The first restriction is sufficient for avoiding inconsistency.[14] The second restriction is in place to secure compositionality.[15]

To see why the second restriction is needed for that job, let us contrast the case of B4 with the example $T(c_\phi) \vee \neg T(c_\phi)$ where c_ϕ denotes some sentence ϕ which is not $T(c_\phi) \vee \neg T(c_\phi)$ itself or either of its disjuncts. A tree for this disjunction is the one shown in Figure 6.2. Assigning values to the nodes

[14] Theorem 6.6 below does not depend on the second restriction.

[15] So if we were happy with true disjunctions without true disjuncts (as in Kripke's supervaluation), we could formulate a version of the theory with just the first restriction and handle Gupta's Challenge (unlike in Kripke's supervaluation) with that alone.

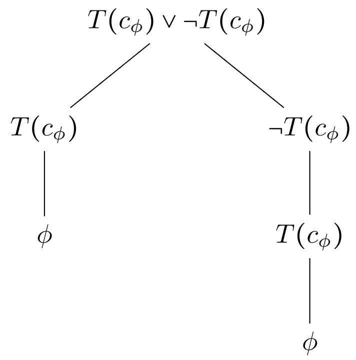


Figure 6.2

of this tree in the same way as before, we again arrive at three possibilities. We can assign $\top$ to the end nodes, which results in the root also getting the value $\top$. Assigning $\perp$ to the end nodes has the same result. When the end nodes are given the value +, so must the nodes above them including the root, for otherwise we would make any such disjunction true even though there is no guarantee that either of the disjuncts will become true.

As will be demonstrated, this method of supervaluation is "reluctant" in assigning truth and falsity, to the point that if it does, the compositional demands for that truth value are guaranteed to be satisfied. For example, if a disjunction is made true, then one of its disjuncts will be too.

The theory to be formulated will be a modification of Kripke's. We adopt the technique of reaching a fixed point through a transfinite series of levels of increasingly more extensive partial interpretations of the truth predicate, using a jump rule to get from one level to the next, and taking unions at limit levels. Only the jump rule is new. The assignment of a truth value to a given sentence at a given level is now decided by considering trees for the sentence.

A tree for a given sentence is constructed by placing that sentence at the root, and below that placing nodes with the *constituents* of that sentence, and then iterating. By "constituents" I mean immediate syntactical parts, in the case of connectives; all the instances, in the case of quantifiers; and the sentence referred to, in the case of a sentence claiming the truth of another sentence. The "treebuilding" is stopped when an atomic sentence with an ordinary predicate is reached, when the root sentence reappears, before any other sentence appears so as to have itself as predecessor (in order not to make it a revision theory) or at any earlier point. Also, we only consider trees for which it holds that each "route" from the root down through the tree is finite, so that truth values can be assigned to the nodes in a well-founded way and "travel" all the way from the end nodes to the root.

A tree is evaluated in the direction from the end nodes to the root. The values of some nodes are fixed: if the sentence at a given node has a truth value from an earlier level, the node has that value; if it is an atomic sentence with an ordinary predicate, its value is decided in the normal way; if it is an atomic sentence with the truth predicate and a constant not denoting a sentence, it is false. The remaining end nodes can have any of the values true, false and undefined, albeit with the restriction that if two nodes are labeled with the same sentence, they must have the same truth value. Indeed, we impose this demand not only on the end nodes, but on all nodes except for the root.[16] The remaining non-end nodes are given values based on the values of the nodes immediately below them. For this, the strong Kleene scheme and Tarski's T-schema with the previously explained modification are used.

In this way, the three problems mentioned at the end of the last section are solved. First, the fixed point is compositional. Second, the three-valued evaluations are reached by quantifying over three-valued evaluations, such that we are genuinely taking all possibilities into account. And third, this theory is not vulnerable to iterated uses of the truth predicate in versions of Gupta's Challenge: if A3* is inserted between the root and the A3 node in the tree in Figure 6.1, and/or A5* is inserted similarly, the result is the same; the truth values merely have to "travel up" one more node. With this theory we check the consequences, for the truth value of a sentence, of the assignment of different truth values to other sentences through several iterations of the truth predicate, while Kripke's theory arbitrarily restricts such consequence-checking to just one iteration.

6.6 The formal theory

We are now ready to formulate the theory precisely, beginning with a rather long list of definitions relating to trees. A **tree**[17] is a triple $Tr = (N, <, l)$ such that N is any set; $<$ is a partial order on N such that for every element of N, the set of predecessors of this element is linearly ordered and finite and there is an element of N called the **root** that is a predecessor of every other element of N; and l is a function from N to $\mathcal{S}$. The elements of N are called **nodes**; a node without successors is called an **end node**; and for each node n, $l(n)$ is

[16]The grounding relation has two relata: that which is grounded in something and that which it is grounded in. The root represents the former relata and the rest of the nodes the latter relata. That is why we should not extend the restriction to also involve the root. Doing so would have the consequence that a sentence would be made true simply because truth is the only proper truth value it can consistently have. And it would not be groundedness-acceptable to have that as a sufficient condition.

[17]This definition is more restrictive than what is standard.

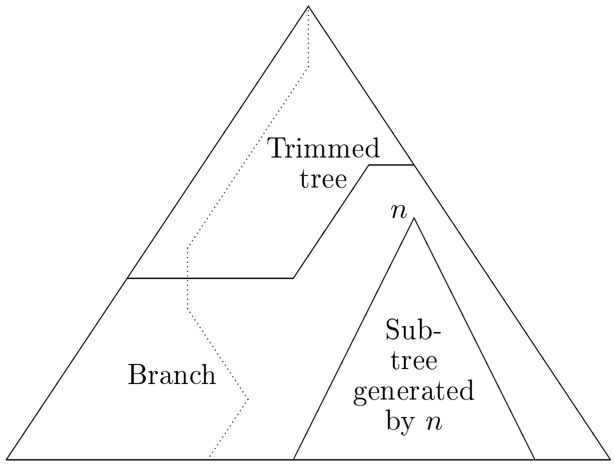


Figure 6.3

called the **label of** n. If a node n has a unique immediate successor labeled ϕ, then "n_ϕ" denotes this successor.

We will consider isomorphic trees to be identical. So, what the elements of N are is of no importance – only the cardinality of N is. (Just think of them as the dots that are connected in a typical graphical representation of a tree.)

Given a tree $Tr = (N, <, l)$, a **trimmed tree of** Tr is defined to be a triple $Tr' = (N', <', l')$ in which N' is a subset of N, such that

- for each node n of N, if $n \in N'$, then all the predecessors of n (by $<$) are as well, and
- for each node n of N, of the immediate successors of n, either all of them or none of them are in N',

and $<'$ and l' are the restrictions of $<$ and l, respectively, to N'. This and the next two definitions are illustrated in Figure 6.3.

Given a tree $Tr = (N, <, l)$ and a node $n \in N$, **the sub-tree of** Tr **generated by** n is the triple $Tr' = (N', <', l')$ such that N' is the subset of N consisting of n and all its successors, and $<'$ and l' are again the restrictions of $<$ and l, respectively, to N'. Note that both trimmed trees and sub-trees generated by a node are trees.

A **branch of** Tr is a maximal linearly ordered subset of the nodes of Tr.

Given a sentence ξ, the **constituents of** ξ are

- ϕ if ξ is $\neg\phi$,
- ϕ and ψ if ξ is $\phi \vee \psi$,

- every sentence of the form $\phi(v/c)$ where c is a constant if ξ is $\exists v\phi$,
- the sentence $I(c)$ if ξ is $T(c)$ and $I(c)$ is a sentence,
- nothing if ξ is $T(c)$ and $I(c)$ is not a sentence, and
- nothing if ξ is $P(c_1, \ldots, c_n)$ where P is an ordinary predicate.

Given a sentence ξ, **the full tree for** ξ is the tree such that the root is labeled with ξ and for every node n, the following holds: 1) If $l(n)$ is ξ and n is not the root, or one of the constituents of $l(n)$ is the label of a non-root predecessor of n, then n has no successors. 2) Otherwise n has one immediate successor for each of the constituents of $l(n)$, and these successors are labeled with these constituents. A trimmed tree of the full tree for ξ is called **a tree for** ξ if each branch thereof is finite.

Note that this means that the full tree for ξ is not *a* tree for ξ if it has infinite branches, as is the case for, e.g., the sentences of Yablo's (1993) Paradox. To be able to evaluate a tree it has to bottom out in end nodes and therefore we cannot always use the full tree. We therefore have to cut it of at some point. This introduces a complication: there is no canonical way to decide how far down to cut it of, so we need to consider a set of trees for each sentence instead of just one. The complication is small though, for if one tree is adequate to rule out enough combinations of truth values to make ξ true, say, then those combinations really are impossible and ξ should be made true. On the other hand, if a given combination is not ruled out, that may just be because the tree is too small. (Consider for example the tree in Figure 6.1 with the bottom row of B nodes removed; using that trimmed tree it would not be possible to rule out that both A3 and A5 could be true.) Therefore, it is reasonable to say that a sentence is given a truth value when just one tree provides that judgment, i.e. when all the evaluations for that tree agree.

We define an **evaluation of a tree for some sentence ξ relative to some evaluation $\mathcal{E} = (T, F)$** (a "tree-evaluation" or, when there is no risk of misunderstanding, just "evaluation") as a function e from the nodes of the tree to $\{\top, \bot, +\}$ such that for every node n,

SEa) if n is not the root, then for every other non-root node n' such that $l(n) = l(n')$, $e(n)$ is identical to $e(n')$;

SEb) if $l(n)$ is in T (F) and n is not the root, then $e(n)$ equals $\top$ ($\bot$);

SEc) if $l(n)$ is of the form $T(c)$ where c is a constant and $I(c)$ is not a sentence, then $e(n) = \bot$; and

SE1) if $l(n)$ is of the form $P(c_1, \ldots, c_n)$ where P is an ordinary n-ary predicate and $c_1, \ldots, c_n$ are constants, then

- $e(n) = \top$ if $(I(c_1), \ldots, I(c_n)) \in I(P)$, and
- $e(n) = \bot$ otherwise,

and for every non-end node n, the following is true:

SE2) If $l(n)$ is of the form $\neg\phi$ where ϕ is a sentence, then

- $e(n) = \top$ if $e(n_\phi) = \bot$,
- $e(n) = \bot$ if $e(n_\phi) = \top$, and
- $e(n) = +$ otherwise.

SE3) If $l(n)$ is of the form $\phi \vee \psi$ where ϕ and ψ are sentences, then

- $e(n) = \top$ if $e(n_\phi) = \top$ or $e(n_\psi) = \top$,
- $e(n) = \bot$ if $e(n_\phi) = \bot$ and $e(n_\psi) = \bot$, and
- $e(n) = +$ otherwise.

SE4) If $l(n)$ is of the form $\exists x\phi$ where x is a variable and ϕ is a formula with at most x free, then

- $e(n) = \top$ if there exists a constant c such that $e(n_{\phi(x/c)}) = \top$,
- $e(n) = \bot$ if for every constant c it holds that $e(n_{\phi(x/c)}) = \bot$, and
- $e(n) = +$ otherwise.

SE5) If $l(n)$ is of the form $T(c)$ where c is a constant, then

- $e(n) = \top$ if $e(n_{I(c)}) = \top$,
- $e(n) = \bot$ if $e(n_{I(c)}) = \bot$,
- $e(n) = \bot$ if $e(n_{I(c)}) = +$ and there is a node labeled $l(n)$ with a successor m with $l(m) = \xi$ and $e(m) = +$, and
- $e(n) = +$ otherwise.

The third bullet of SE5 is the implementation of the strong truth rule. The second conjunct of the condition imposes the restriction on the use of that strong rule explained in the italicized sentence on page 163, above.

We define the **supervaluation with respect to the evaluation** $\mathcal{E} = (T, F)$ as $\mathrm{SE}_\mathcal{E} = (ST_\mathcal{E}, SF_\mathcal{E})$ where $ST_\mathcal{E}$ ($SF_\mathcal{E}$) is the set of those sentences ξ, such that for some tree for ξ, all evaluations of this tree relative to $\mathcal{E}$ have $\top$ ($\bot$) as the value of the root.[18] Such a tree **decides** ξ **with respect to** $\mathcal{E}$, and we

[18] Note that if there is no evaluation of some tree for a sentence, the sentence becomes both true and false. It needs to be proved that this situation cannot arise.

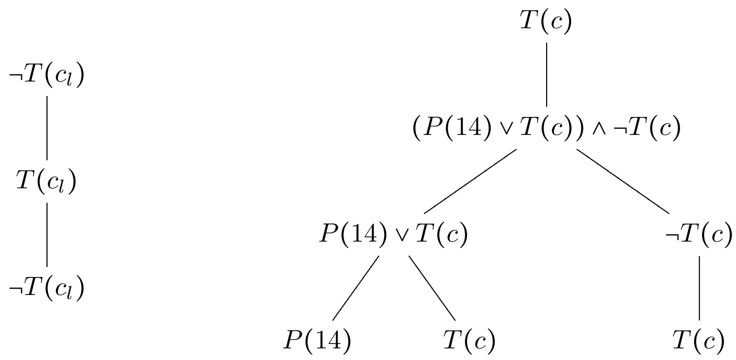


Figure 6.4

say that $\mathcal{E}$ **makes** ξ **true (false)**. In contrast, saying that ξ **is true (false)** ***in*** $\mathcal{E}$ still means that $\xi \in T$ ($\xi \in F$).

For all ordinals α, the **supervaluation at level** α, SE^α, is defined by recursion:

$$\mathrm{SE}^\alpha = \begin{cases} (\varnothing, \varnothing) & \text{if } \alpha = 0 \\ \mathrm{SE}_{\mathrm{SE}^{\alpha-1}} & \text{if } \alpha \text{ is a successor ordinal} \\ \left(\bigcup_{\eta<\alpha} ST_{\mathrm{SE}^\eta}, \bigcup_{\eta<\alpha} SF_{\mathrm{SE}^\eta}\right) & \text{if } \alpha \text{ is a limit ordinal} \neq 0 \end{cases}$$

A tree **decides a sentence at a successor level** α, if the tree decides the sentence with respect to $\mathrm{SE}^{\alpha-1}$, and the sentence is then **made true/false at level** α.

Figure 6.4 shows two examples of trees. The example on the left is the full tree, and also *a* tree, for the Liar. That is, c_l is a constant denoting $\neg T(c_l)$. At the first level, this tree has three evaluations. The first assigns $\top$ to the end node and to the "middle node" and $\bot$ to the root. The second is the other way around: $\bot$ to the two bottom nodes and $\top$ to the root. The third assigns $+$ to the end node. Then the third bullet of SE5 kicks in and assigns $\bot$ to the middle node. So again the root is assigned $\top$. Ergo, the evaluations do not agree on a value for the root and hence $\neg T(c_l)$ is not given a truth value in the supervaluation. As the reader can easily verify, the smaller trees for $\neg T(c_l)$, along with the trees for $T(c_l)$, also have disagreeing evaluations. Therefore, the evaluations for these trees are exactly the same at every level.

The example on the right is about the sentence $T(c)$ where $I(c) = (P(14) \vee T(c)) \wedge \neg T(c)$. The predicate P means "is prime" and 14 is a constant with the obvious denotation. Intuitively $P(14) \vee T(c)$ and $\neg T(c)$ contradict each other, since 14 is composite. Ergo $T(c)$ should be false. As in the Liar example, the tree depicted here is both the full tree and *a* tree for the root sentence. And at the first level this tree also has three evaluations; SEa and SE1 prevent there from being more. They all assign $\bot$ to the node labeled $P(14)$. The first further assigns $\top$ to the end nodes labeled $T(c)$ and to the node labeled $P(14) \vee T(c)$ and $\bot$ to all the remaining nodes. The second makes those end nodes $\bot$, the node labeled $\neg T(c)$ $\top$, and the rest of the nodes $\bot$. The third leaves the two end nodes with a +, resulting in the three nodes in the middle getting + too, whereupon the third bullet of SE5 again takes effect so that the root is assigned $\bot$. In this case, all the evaluations agree on the root, so $T(c)$ is made false at level 1. As a consequence, at all higher levels this tree only has one evaluation, namely the second mentioned of the three. At level 2, $\neg T(c_\phi)$ is made true by the two-node sub-tree generated by the node labeled with that sentence. Similarly, $P(14) \vee T(c)$ and $(P(14) \vee T(c)) \wedge \neg T(c)$ are made false at level 2, securing compositionality.

In order to conclude the statement of the theory we need to prove monotonicity, the existence of a fixed point, and consistency, but this is postponed to Section 6.9. We refer to the fixed point as $\mathcal{SE}$. For all sentences ξ we say that ξ is **true** if ξ is in the truth set of $\mathcal{SE}$, **false** if ξ is in the falsity set of $\mathcal{SE}$, and **undefined** otherwise.

6.7 Meeting Gupta's Challenge

We can now apply the theory to Gupta's Challenge. Let A and B be unary predicates interpreted as "is a sentence spoken by A" and "is a sentence spoken by B" respectively; and let = be a binary predicate meaning "are identical". Then the sentences A3, A5 and B4 can be formalized as follows:

(A3) $$\forall x(B(x) \to T(x))$$

(A5) $$\exists x(B(x) \wedge \neg T(x))$$

(B4) $$\forall x \forall y(A(x) \wedge T(x) \wedge A(y) \wedge T(y) \to =(x,y))$$

A tree for (B4) is outlined in Figure 6.5. The constants c_{A3}, c_{A5}, and c_{B4} refer to (A3), (A5), and (B4) respectively. It is here assumed for the sake of simplicity that no other constants refer to these sentences. Another simplification is that it is pretended that conjunction and the conditional are primitive connectives and that multiple sentences can be concatenated with connectives going just one node up. I leave it to the reader to verify that these simplifications do not affect the conclusion.

It is easily verified that all instances of the doubly universally quantified sentence (B4) which are not displayed in the figure are assigned the value $\top$. Therefore there are only three essentially different evaluations of this tree.

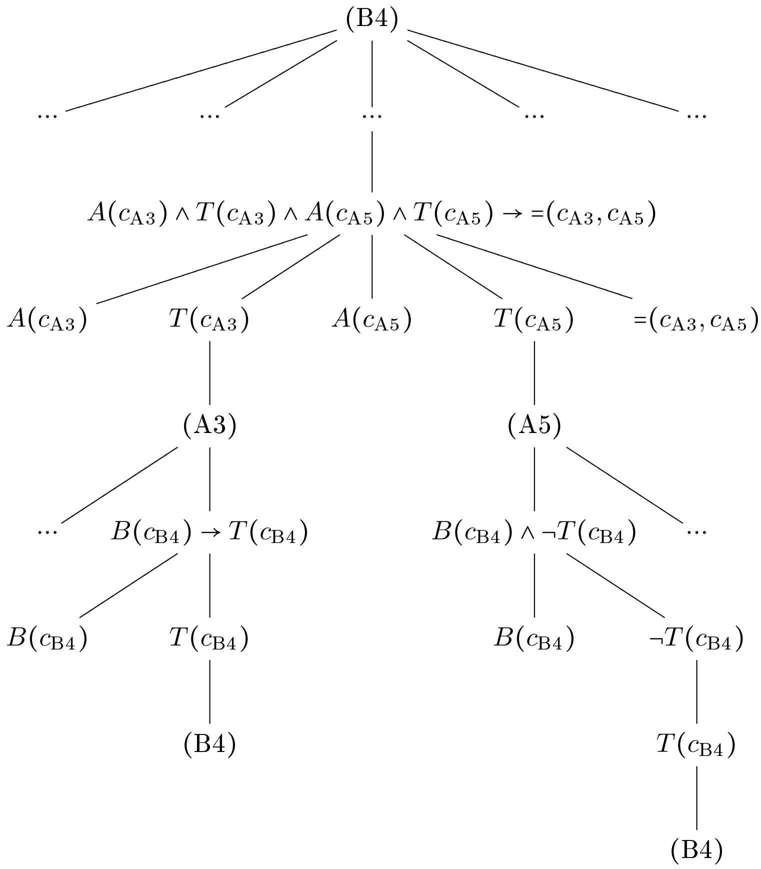


Figure 6.5

First, there is a tree-evaluation that gives the value $\top$ to the two end nodes labeled (B4). In this evaluation, the nodes labeled $T(c_{B4})$ have the value $\top$, and the node labeled $\neg T(c_{B4})$ the value $\bot$. The nodes with $B(c_{B4})$ have the value $\top$. Ergo, the nodes labeled $B(c_{B4}) \rightarrow T(c_{B4})$ and $B(c_{B4}) \wedge \neg T(c_{B4})$ have the values $\top$ and $\bot$ respectively. Under the node labeled (A3) there is an infinity of nodes, of which all that are not shown also have the value $\top$, so this node also has the value $\top$ in this evaluation. Similarly, all the other nodes under the node labeled (A5) have the value $\bot$, so this node does as well. It follows that the two nodes labeled $T(c_{A3})$ and $T(c_{A5})$ have the values $\top$ and $\bot$ respectively. Hence, the node labeled $A(c_{A3}) \wedge T(c_{A3}) \wedge A(c_{A5}) \wedge T(c_{A5}) \rightarrow =(c_{A3}, c_{A5})$ has the value $\top$. All other nodes with sentences of the same form also have the value $\top$, and therefore the root must too.

Second, there is a tree-evaluation that assigns the value $\bot$ to the two end nodes labeled (B4). The fact that this evaluation also makes the root true follows by analogous reasoning.

The third and final tree-evaluation gives the value + to the two nodes labeled (B4). Then, by the third bullet of SE5, the nodes labeled $T(c_{B4})$ have the value $\perp$, and therefore the rest of the evaluation is exactly as when $\perp$ was assigned to the two end nodes labeled (B4).

Because of SEa, there are no other tree-evaluations. From this, it follows from the definition of supervaluation that (B4) is made true at level 1. From monotonicity it then follows that (B4) is true.

Then at level 2, we only need to look at the tree for (A3) which ends at the node labeled (B4) (or to put it more precisely: the tree for (A3) which is the sub-tree generated by the node labeled (A3) in the tree in the figure). Now there is only one evaluation of this tree, namely the evaluation in which the node labeled (B4), and hence the root, is given the value $\top$. That is, (A3) is determined to be true at level 2. Similarly and simultaneously, (A5) is made false.

So this theory handles Gupta's Challenge as desired. And Gupta's "revenge" in the form of replacing A3 and A5 with A3* and A5* does not bite. The proof goes through with minor modifications: Insert two more nodes with labels $T(c_{TA3})$ and $T(c_{TA5})$ in the middle of the tree as immediate predecessors of the nodes labeled $T(c_{A3})$ and $T(c_{A5})$, respectively. The constant c_{TA3} obviously refers to $T(c_{A3})$, and c_{TA5} to $T(c_{A5})$. The three evaluations at level 1 are essentially as before; the truth values just have to "travel up one more node" with $T(c_{TA3})$ getting the same truth value as $T(c_{A3})$, and $T(c_{TA5})$ getting the same the same as $T(c_{A5})$. The combination of A3 with A5* works just as well.

6.8 A generalization

However, the theory, as stated, does not handle all Gupta-style challenges adequately. If we modify the story, so that person B says one more sentence, namely

B5: At least two things A says are true

intuition still insists that all ten sentences have proper truth values. The reasoning is only a slight modification of that in Section 6.1: A3 and A5 contradict each other, so at most one of them can be true. Hence at most one thing A says is true. So B4 is true and B5 is false. Ergo, A3 is false and A5 is true. These truth values are also, to use the language of Section 6.2, grounded in facts about which combinations of truth values are possible.

Figure 6.6 shows, in simplified form, trees for B4 and B5. The following tables detail the nine evaluations of each of the trees.

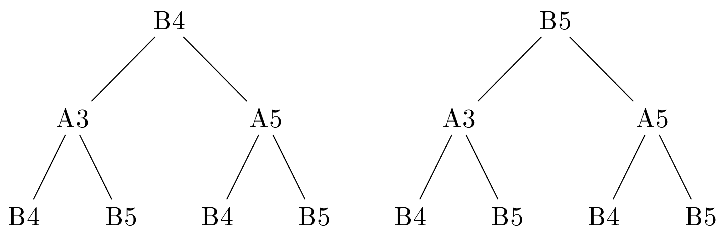


Figure 6.6

B4 (root)	⊤	+	⊤	⊤	⊤	⊤	⊤	⊤	⊤
A3	⊤	+	⊥	⊥	⊥	⊥	⊥	⊥	⊥
A5	⊥	+	⊤	⊤	⊤	⊤	⊤	⊤	⊤
B4	⊤	⊤	⊤	+	+	+	⊥	⊥	⊥
B5	⊤	+	⊥	⊤	+	⊥	⊤	+	⊥

B5 (root)	⊥	⊥	⊥	+	⊥	⊥	⊥	⊥	⊥
A3	⊤	⊥	⊥	+	⊥	⊥	⊥	⊥	⊥
A5	⊥	⊤	⊤	+	⊤	⊤	⊤	⊤	⊤
B4	⊤	⊤	⊤	+	+	+	⊥	⊥	⊥
B5	⊤	+	⊥	⊤	+	⊥	⊤	+	⊥

The top row of the first table only contains the values true and undefined and so "almost" makes B4 true. Not quite, though, for in one tree-evaluation the root is undefined. Similarly, B5 is not supervaluated as false.

Making B4 true and B5 false can be done without taking the risk of applying the strong rule for the truth predicate to an undefined sentence that at a later stage becomes true. We just need to consider the two sentences "simultaneously", a complication that the present rules do not take into account.

We can remedy the situation by amending the theory as follows: For each set of sentences S, we consider all sets of trees containing exactly one tree Tr_ξ for each $\xi \in S$, such that elements of S are only labels of root nodes and end nodes. Then we replace "$l(m) = \xi$" in the third bullet of SE5 with "$l(m) \in S$". If it holds for each Tr_ξ that all evaluations of it assign the same value to the root, then all the sentences in S are given these root values. With this change, which is a straightforward generalization of the original theory in which S was only allowed to be a singleton, B4 is made true and B5 false, in accordance with the intuitive verdict. A3 and A5 are similarly given the right truth values (now already at level 1).

To be more precise, we change four of the definitions. First, the definition of full tree is altered to read as follows: given a sentence ξ *and a set of sentences S containing ξ*, **the full tree for ξ *relative to* S** is the tree such that the root is labeled with ξ and for every node n, the following holds: 1) If $l(n)$ is *an element of* S and n is not the root, or one of the constituents of $l(n)$ is the label of a non-root predecessor of n, then n has no successors. 2) Otherwise, n has one immediate successor for each of the constituents of $l(n)$, and these successors are labeled with these constituents. For some examples, consider Figure 6.4 again. The tree on the left is the full tree for $\neg T(c_l)$ relative to S for any S that does not include $T(c_l)$. If S does include $T(c_l)$, the full tree stops one node higher up. Similarly, the tree on the right is the full tree for $T(c)$ relative to any S that does not include any of the three sentences $(P(14) \vee T(c)) \wedge \neg T(c)$, $P(14) \vee T(c)$ and $\neg T(c)$. If it does include, say, $P(14) \vee T(c)$, remove the two left-most end nodes to get the full tree for $T(c)$ relative to S.

Second, the definition of a tree is relativized to S in the obvious way. Third, the definition of tree-evaluation is relativized to S by changing "$l(m) = \xi$" to "$l(m) \in S$". And fourth, **supervaluation with respect to the evaluation** $\mathcal{E} = (T, F)$ is redefined as $\mathrm{SE}_{\mathcal{E}} = (ST_{\mathcal{E}}, SF_{\mathcal{E}})$ where $ST_{\mathcal{E}}$ ($SF_{\mathcal{E}}$) is the set of those sentences ξ, such that for *some set of sentences S containing ξ and* some tree for ξ *relative to S*, all evaluations of this tree relative to $\mathcal{E}$ *and S* has $\top$ ($\bot$) as the value of the root*; and further, that for all the other elements of S there are also trees for them relative to S, all the evaluations of which agree on assigning $\top$ or agree on assigning $\bot$ to the root.* Such a tree **decides ξ with respect to $\mathcal{E}$**, and we say that $\mathcal{E}$ **makes ξ true (false)** ***(among S)***.

6.9 Theorems and proofs

For this generalized theory we can now prove the promised theorems.

Lemma 6.1. *For any evaluation $\mathcal{E}$, extension $\mathcal{E}'$ of $\mathcal{E}$, sentence ξ, set of sentences S containing ξ, tree Tr for ξ relative to S, and evaluation e of Tr relative to S and $\mathcal{E}'$, the tree-evaluation e is also an evaluation of Tr relative to S and $\mathcal{E}$.*

Proof. If one of the antecedents of SEb is satisfied for $\mathcal{E}$, then the same antecedent is satisfied for $\mathcal{E}'$. So a restriction on what counts as an evaluation of the tree imposed by this clause in the case of $\mathcal{E}$ also applies in the case of $\mathcal{E}'$. The same holds trivially for the other clauses. □

Lemma 6.2. *For any evaluation $\mathcal{E}$ and extension $\mathcal{E}'$ of $\mathcal{E}$, $\mathrm{SE}_{\mathcal{E}'}$ is an extension of $\mathrm{SE}_{\mathcal{E}}$.*

Proof. For any sentence ξ, set of sentences S containing ξ, and tree for that sentence relative to S, the set of evaluations of the tree relative to S and $\mathcal{E}'$ is a subset of the set of evaluations of the tree relative to $\mathcal{E}$. This follows from Lemma 6.1. So ξ satisfies the criterion for being in $ST_{\mathcal{E}'}$ ($SF_{\mathcal{E}'}$) if it satisfies the criterion for being in $ST_{\mathcal{E}}$ ($SF_{\mathcal{E}}$). □

Theorem 6.3 (Monotonicity). *For all ordinals α and β, if $\alpha < \beta$ then* SE^{β} *is an extension of* SE^{α}.

Proof. As SE^0 is empty, this follows from Lemma 6.2 by induction. □

Theorem 6.4 (Fixed point). *There is an ordinal α such that for all ordinals $\beta > \alpha$,* $\mathrm{SE}^{\beta} = \mathrm{SE}^{\alpha}$.

Proof. This follows from Theorem 6.3 by the usual cardinality argument. □

As prematurely mentioned, we refer to the fixed point as $\mathcal{SE}$.

The basic idea for the consistency proof is to show that the largest intrinsic fixed point of the strong Kleene jump is an extension of $\mathcal{SE}$. For this, we need some definitions. "Fixed point of the strong Kleene jump" will be shortened to **FPSK** (and following Kripke, we will take these to include only consistent evaluations). An evaluation $\mathcal{E}$ is an **intrinsic FPSK** if $\mathcal{E}$ is an FPSK and it is the case that for any FPSK $\mathcal{E}'$ there exists an FPSK $\mathcal{E}''$ that is an extension of both $\mathcal{E}$ and $\mathcal{E}'$. In that case, the elements of the truth (falsity) set of $\mathcal{E}$ are called **intrinsically true (false)**. Of the intrinsic FPSKs, one is the largest (Kripke 1975, 709). We denote it $\mathcal{I}$.

For any set of sentences S, function $\pi : S \to \{\top, \bot\}$ and FPSK $\mathcal{E} = (T, F)$, let $\mathcal{E}^{\pi}$ be $(T \cup \pi^{-1}(\top), F \cup \pi^{-1}(\bot))$. The **closure of** $(\mathcal{E}, \pi)$, denoted $cl(\mathcal{E}, \pi)$, is defined as the smallest extension (CT, CF) of $\mathcal{E}^{\pi}$ such that if $\phi \in CT$ or $\psi \in CT$, then $\phi \vee \psi \in CT$; and if $\phi \in CF$ and $\psi \in CF$ then $\phi \vee \psi \in CF$ and similarly for negation, the existential quantifier and the truth predicate (we refer to these as **closure rules**). The **semi-closure of** $(\mathcal{E}, \pi)$, denoted $cl^{-}(\mathcal{E}, \pi)$, is defined as the smallest extension (CT, CF) of $\mathcal{E}^{\pi}$ such that if $\phi \in CT$ or $\psi \in CT$ *and* $\phi \vee \psi \notin S$, then $\phi \vee \psi \in CT$; and if $\phi \in CF$ and $\psi \in CF$ *and* $\phi \vee \psi \notin S$, then $\phi \vee \psi \in CF$ and similarly for negation, etc.

There are two things about these definitions that should be noted. First, closure could just as well have been defined as a function on $\mathcal{E}^{\pi}$, for it does not matter which truths and falsities "comes from" $\mathcal{E}$ and which from π. The only reason for not defining it like that is the desire to have the wording of the definition as close as possible to that of semi-closure. For semi-closure it *does* matter what comes from $\mathcal{E}$ and what from π. Second, $cl(\mathcal{E}, \pi)$ may not be an FPSK. As no restrictions have been placed on π, it may, for example, take a disjunction to $\top$, so that that disjunction is also true in $cl(\mathcal{E}, \pi)$, without either of the disjuncts being true in $cl(\mathcal{E}, \pi)$. Another reason that $cl(\mathcal{E}, \pi)$ may not be an FPSK is that it can be inconsistent. Taking the closure of $(\mathcal{E}, \pi)$ is like doing the Kripke iteration starting from $\mathcal{E}^{\pi}$ except that monotonicity is forced. The following lemma gives a condition under which $cl(\mathcal{E}, \pi)$ *is* an FPSK:

Lemma 6.5. *Let S be a set of sentences, π a function $S \to \{\top, \bot\}$ and $\mathcal{E}$ an FPSK such that $cl(\mathcal{E}, \pi)$ is consistent. For each $\xi \in S$, let $Tr_\xi = (N_\xi, <_\xi, l_\xi)$ be a tree for ξ relative to S with more than one node. Let e_ξ be the function from N_ξ to $\{\top, \bot, +\}$ defined by having, for each $n \in N_\xi$, $e(n) = \top$ ($\bot$; $+$) if $l(n)$ is true (false; undefined) in $cl(\mathcal{E}, \pi)$. If for each ξ, e_ξ is an evaluation of Tr_ξ relative to S and $(\varnothing, \varnothing)$, then $cl(\mathcal{E}, \pi)$ is an FPSK.*

Proof. Let $\mathcal{E}^*$ be the result of applying the strong Kleene jump to $cl(\mathcal{E}, \pi)$. We need to show that $cl(\mathcal{E}, \pi)$ is an extension of $\mathcal{E}^*$ and that $\mathcal{E}^*$ is an extension of $cl(\mathcal{E}, \pi)$. The former follows directly from the definition of cl. To demonstrate the latter it is enough to show that for all $\xi \in S$, ξ is false in $\mathcal{E}^*$ if $\pi(\xi) = \bot$ and true in $\mathcal{E}^*$ if $\pi(\xi) = \top$. So let an $\xi \in S$ be given.

As e_ξ is an evaluation of Tr_ξ and assigns $\pi(\xi)$ to the root labeled ξ, e_ξ and thereby $cl(\mathcal{E}, \pi)$ assign values to the immediate successors of the root/the constituents of ξ in such a way that (if ξ is an atomic sentence with the truth predicate) the T-scheme is satisfied (because the assignment of $\pi(\xi')$ to end nodes labeled with a $\xi' \in S$ implies that the third bullet of SE5 is not applied) or (if ξ is any other type of sentence) the strong Kleene scheme is satisfied.

Each of those constituents ϕ satisfies the T-schema/the strong Kleene scheme in relation to *its* constituents: if ϕ is an element of S, this follows from similar considerations about the root and its immediate successors in Tr_ϕ, and if not, it follows from the definition of $cl(\mathcal{E}, \pi)$.

For each constituent, continue like this until either 1) reaching sentences that do not have constituents or 2) passing from atomic sentences with the truth predicate to their constituents. Having the latter (with their associated truth values) in $cl(\mathcal{E}, \pi)$ is sufficient for having ξ false in $\mathcal{E}^*$ if $\pi(\xi) = \bot$ and having ξ true in $\mathcal{E}^*$ if $\pi(\xi) = \top$. □

Theorem 6.6. *$\mathcal{SE}$ is consistent.*

Proof. We prove by induction that SE^α is consistent and that $\mathcal{I}$ is an extension of SE^α. The base and limit cases are trivial. So for the successor case, let an α be given and assume that $\mathcal{I}$ is an extension of SE^α. Let ξ be a sentence.

We first prove that ξ is not made both true and false by SE^α among the same S. To do so it must be demonstrated that a) every tree for ξ relative to S has an evaluation relative to SE^α and b) there is not one tree for ξ relative to S all the evaluations for which assign $\top$ to the root, and another tree for ξ relative to S all the evaluations for which assign $\bot$ to the root.

Let Tr_ξ be a tree for ξ relative to S. We construct an evaluation for Tr_ξ as follows:[19]

1) For each node n of Tr_ξ, if $l(n)$ is true (false) in $\mathcal{I}$, then assign the value $\top$ ($\bot$) to n.
2) If there are end nodes labeled with an element of S that were not assigned a value in step 1, assign them values: $\top$ if the label is made true by SE$^\alpha$, $\bot$ otherwise.
3) Starting from those end nodes labeled with an element of S and going through their predecessors from the bottom up, ensure that the first two bullets of each of SE2–SE5 are satisfied.
4) For each non-end node that is assigned a value in step 3, if there are other non-end nodes with the same label, copy the assigned value to them. Then repeat steps 3 and 4 starting from those nodes.
5) Assign + to all remaining nodes.

(To prevent confusion: In fact no node is assigned $\top$ in step 2. But that will only have been established at the end of this proof. It does not follow *directly* from the induction hypothesis, for being made true *by* SE$^\alpha$ is not the same as being true *in* SE$^\alpha$.)

To see that this is an evaluation we first need to verify that every node is assigned exactly one of the three values. As step 5 assigns + to all nodes that have not already received a truth value, and + is assigned in no other step, this reduces to showing that no node is assigned both $\top$ and $\bot$ in the first four steps. The consistency of $\mathcal{I}$ implies that this does not happen in step 1, and step 2 expressly avoids nodes that have been given a value in step 1 and at most assigns one value to other nodes. For steps 3 and 4 I will just give the intuitive idea for what should formally be an induction argument concerning a transfinite, monotonic sequence of partial tree-evaluations where values are

[19] The most straightforward approach to constructing such an evaluation will not work. If we simply assign values to the end nodes according to the rule "$\top$ if the label is in the truth set of SE$^\alpha$/$\bot$ if the label is in the falsity set/+ otherwise", and then assign values to the non-end nodes according to SE2-SE5, the result is not necessarily a tree-evaluation relative to the given evaluation. For example, let ϕ and ψ be labels of two end nodes and assume they are in neither the truth set nor the falsity set. Just above them may be a node labeled $\phi \vee \psi$ which *is* in the truth set. According to SEb, the node should be assigned $\top$, while SE3 dictates that it should instead be given the value +. (As is proved below, either ϕ or ψ would eventually be made true, but compositionality is something that holds in the fixed point, not at every level leading up to it.)

added to nodes one at a time (in steps 3 and 4 from the bottom up), and I will stick to the example of a node n that is labeled with a disjunction $\phi \vee \psi$ and assigned $\bot$ at some point in the "process", leaving the other cases to the reader. We need to show that this node was not already assigned $\top$ previously in the process, and we can assume that at this point no other node is assigned both $\top$ and $\bot$.

If n is assigned $\bot$ in step 3, it has successors labeled ϕ and ψ that are both assigned $\bot$ and therefore not $\top$. Hence, those two sentences are not true in $\mathcal{I}$, so neither is $\phi \vee \psi$, ergo n was not assigned $\top$ in step 1. It also follows directly from those two nodes not having been assigned $\top$ that n has not been assigned $\top$ in step 3. Since n is a non-end node, it was not assigned any value in step 2. Finally, it can not have been assigned $\top$ in step 4 (even though *a* step 4 can precede *a* step 3), for that would imply that the node from which it was copied had successors labeled ϕ and ψ at least one of which would have been assigned $\top$. That is impossible, for had that assignment happened in step 1 or 2, it would also have happened to the successor of n with the same label (in the latter case because they would then both be end nodes); and had it happened in step 3, it would have been copied in step 4.

If instead n is assigned $\bot$ in step 4, then some other node m labeled $\phi \vee \psi$ was assigned $\bot$ in step 3 and then the same reasoning can be used to show that n was not assigned $\top$ in steps 1, 3 or 4. That it was also not in step 2 is due to the fact that it follows from m being assigned a value in step 3 that m is a non-end node and that therefore $\phi \vee \psi$ is not in S.

We can now proceed to show that what is constructed is an evaluation of Tr_ξ relative to S and SE^α by verifying that all the clauses are satisfied. The induction hypothesis implies that SEb is satisfied after step 1. SEc and SE1 are as well. SEa is satisfied for the following reason: it is obvious that it is after step 1; in step 2 values are only assigned to nodes with a label that only appear on other end nodes, save perhaps the root, so the same values are assigned to all other non-root nodes with the same labels in that step; and step 4 is in place to ensure that violations of SEa in step 3 are taken care of. Bullets one and two of SE2–SE5 are satisfied after step 1, may not be after step 2, but then are again after step 3. Let m be a node that has a value "copied" to it in step 4, and assume that m has immediate successors (there is nothing to show if not). The node from which the value was copied has immediate successors with the same labels, and there the first two bullets of SE2–SE5 were satisfied, so they are also satisfied for m and its immediate successors. Step 2 ensures that the third bullet of SE5 is vacuously satisfied. The last bullet of each of SE2–SE5 is satisfied by step 5.

With that, part a) is dealt with. For part b), let Tr'_ξ and Tr''_ξ be two trees for ξ relative to S. Let Tr_ξ be their "union" (in the obvious but not literal meaning of this word). Tr_ξ is also a tree for ξ relative to S, as the properties of having only finite branches and only having the root and end nodes labeled with elements of S are preserved. It is easily seen that, excluding evaluations which assign + to end-nodes labeled with elements of S, any evaluation of Tr_ξ can be restricted to an evaluation of Tr'_ξ and an evaluation of Tr''_ξ. As it has been shown that Tr_ξ has such an evaluation, Tr'_ξ and Tr''_ξ have evaluations that assign the same value to their roots.

Now assume that ξ is made true among S (the case of falsity is similar). We show that ξ is intrinsically true. Let π be the function that describes the truth values that the elements of S are given. Obviously, ξ is not intrinsically false, for then the constructed tree-evaluation makes the root false, and therefore ξ would not have been made true. By generalization it follows that $\mathcal{I}^\pi$ is consistent. As a closure rule can only produce a truth value for a sentence that contradicts an intrinsic truth or falsity from other truth values that contradict intrinsic truths or falsities, it further follows that $cl^-(\mathcal{I},\pi)$ is also consistent.

Under the assumption that ξ is made true *among* S, what happened in step 2 was that each end node with a label ξ' from S was given exactly the value $\pi(\xi')$. Ergo, for all non-root nodes, the value assigned is the same as the one the label has according to $cl^-(\mathcal{I},\pi)$. From the assumption that ξ is made true, it follows that the root is assigned $\top$; and hence it follows from the fact that the two first bullets of SE2–SE5 are satisfied that ξ cannot be made false by one more application of a closure rule to $cl^-(\mathcal{I},\pi)$. Generalizing this observation to all the elements of S, it follows that $cl(\mathcal{I},\pi) = cl^-(\mathcal{I},\pi)$, i.e. that $cl(\mathcal{I},\pi)$ is consistent. So according to Lemma 6.5, $cl(\mathcal{I},\pi)$ is an FPSK, since the constructed evaluation of Tr_ξ is exactly the function e_ξ mentioned in that lemma.

Assume for *reductio* that ξ is false in some FPSK. By the definition of $\mathcal{I}$ it follows that there is an FPSK, $\mathcal{I}^*$, which is an extension of $\mathcal{I}$ and in which ξ is false. Define a new evaluation of Tr_ξ the same way as above but using $\mathcal{I}^*$ instead of $\mathcal{I}$ in step 1. This is an evaluation of Tr_ξ relative to $\mathcal{I}^*$ and hence according to Lemma 6.1 and the induction hypothesis also relative to SE^α, and it assigns $\bot$ to the root. This is a contradiction.

It can be concluded that ξ is true in an FPSK and not false in any. Hence, no sentence that is true or false in $\mathcal{I}^\pi$ has the opposite truth value in an FPSK. It follows that no sentence that is true or false in $cl(\mathcal{I},\pi)$ has the opposite truth value in an FPSK. Ergo $cl(\mathcal{I},\pi) = \mathcal{I}$, i.e. ξ is intrinsically true. As ξ was arbitrary, $\mathcal{I}$ is an extension of $\mathrm{SE}^{\alpha+1}$. From this it finally follows that ξ is not made both true and false in $\mathrm{SE}^{\alpha+1}$, period. $\square$

Theorem 6.7 ($\mathcal{SE}$ is compositional and satisfies the T-schema)**.** *For all sentences ϕ and ψ, formulae ζ with at most the variable x free, and constants c such that $I(c) \in \mathcal{S}$, the following holds:*

- *$\neg\phi$ is true iff ϕ is false.*
- *$\neg\phi$ is false iff ϕ is true.*
- *$\phi \vee \psi$ is true iff ϕ is true or ψ is true.*
- *$\phi \vee \psi$ is false iff ϕ is false and ψ is false.*
- *$\exists x\zeta$ is true iff for some constant k, $\zeta(x/k)$ is true.*
- *$\exists x\zeta$ is false iff for all constants k, $\zeta(x/k)$ is false.*
- *$T(c)$ is true iff $I(c)$ is true.*
- *$T(c)$ is false iff $I(c)$ is false.*

Proof. The right-to-left direction is, in each case, simple: Consider a level where the right-hand side is satisfied. Then at the next level, the tree for the sentence on the left-hand side consisting of just the root and immediate successors thereof will do the job.

The left-to-right direction can be proved by induction on the smallest level that decides the sentence on the left-hand side. Such a level is always indexed by a successor ordinal, so the base and limit cases are trivial. Let α be an ordinal, and ξ a sentence that is decided at level $\alpha + 1$ and not at any lower level. Let S be a set of sentences that ξ was decided among, and let Tr_ξ be a tree that decided ξ relative to S at level $\alpha + 1$. Again we construct a tree-evaluation, this time by taking these steps:

1) For each node n of Tr_ξ, if $l(n)$ is true (false) in SE^α, then assign the value $\top$ ($\perp$) to n.
2) Starting from each node assigned a value in step 1 and working downwards, ensure that the first two bullets of each of SE2–SE5 are satisfied, by assigning values to immediate successors that their labels have in $\mathcal{SE}$. This is possible according to the induction hypothesis.
3) For each node that is assigned a value in step 2, if there are other nodes with the same label, copy the assigned value to them. Then repeat steps 2 and 3 starting from those nodes.

4) Assign the value of each element of S according to $\mathcal{SE}$ to end nodes labeled with those sentences.

5) Starting from each node assigned a value in steps 1, 3 and 4 and working upwards, ensure that the first two bullets of SE2–SE5 are satisfied.

6) For each node that is assigned a value in step 5, if there are other nodes with the same label, copy the assigned value to them. Then repeat steps 5 and 6 starting from those nodes.

7) Assign + to all remaining nodes.

This is an evaluation of Tr_ξ relative to SE$^\alpha$: SEb, SEc and SE1 are satisfied after step 1. That SEa and SE2–SE5 are satisfied is demonstrated in the same way as in the previous proof.

If follows that this evaluation assigns the same value to the root as ξ has in $\mathcal{SE}$. From this it again follows that, if this value is $\top$ ($\bot$) and ξ is of the form $\neg\phi$, then the immediate successor of the root (labeled ϕ) is assigned $\bot$ ($\top$), and similarly for the other bullets of the theorem; in the last case because step 4 ensures that bullet 3 of SE5 does not come into play. Whenever $\top$ ($\bot$) is assigned to a node, the label of that node is true (false) in $\mathcal{SE}$. This follows from monotonicity in the case of step 1, from the induction hypothesis in the case of step 2, from the right-to-left direction of this proof in the case of step 5, and is trivial for the other steps. From this the desired conclusion can be deduced. □

6.10 Comparison and discussion

Leading up to a discussion of the adequacy of this theory, let us undertake a comparison between the theory presented in this chapter and the various versions of Kripke's theory.

The basic version can be reformulated using trees as follows: For each sentence, consider only the tree for that sentence that stops one node below any node labeled with a sentence of the form $T(c)$ for some constant c. Further, remove the third bullet from SE5, and let SEa and SEb apply only to end nodes. Then check only one evaluation of this tree, namely the one that assigns + to all end nodes to which none of SEb, SEc or SE1 applies. If the root is assigned a proper truth value by this tree-evaluation, then (and only then) is the sentence in question given that value.[20]

[20] The weak Kleene scheme version (which is as the the strong Kleene scheme version except that a sentence is undefined if any constituent of it is) can be obtained similarly if proper modifications are made to clauses SE3 and SE4. Revision theory can be reformulated using

The simple Kripkean supervaluation version can be reconstructed using the same trees and by restricting attention to evaluations that assign proper truth values to all nodes. The more sophisticated supervaluation versions then correspond to considering even fewer different tree-evaluations.

So all these versions of Kripke's theory can be reformulated using trees, and hence, in a sense, the framework of this chapter is a generalization of Kripke's. In fact, they can be reformulated using just a certain class of rather small trees. With the perspective afforded by the more general framework, this can be seen to be an arbitrary and unreasonable restriction. Kripkean supervaluation only looks forward through one iteration of the truth predicate, so to speak. The starred versions of Gupta's Challenge show that a supervaluation method should look forward through an arbitrary number of iterations.

However, there is another sense in which the present theory is not a generalization compared to Kripke's work. He does consider the set of all fixed points for the jump operation in abstraction from methods of "arriving" at these fixed points. And the theory considered in this chapter is just one of these fixed points for the strong Kleene jump. (This follows from Theorem 6.7, together with the fact that atomic sentences with ordinary predicates are assigned truth values the "right" way.) So in that sense, the merit of the tree framework is merely that it provides a method for reaching this fixed point "from below". But that is philosophically important, as it is a necessary condition for it to be groundedness-acceptable.

The proof of Theorem 6.6 reveals that the fixed point is intrinsic, i.e. no sentence has a proper truth value in it if it has the opposite truth value in some other fixed point. However, the fixed point is not the largest intrinsic fixed point, for the following sentence is "intrinsically true" but undefined in the fixed point: "This sentence is true or the Liar is false".[21] This example tells against the largest intrinsic fixed point, not the theory of this chapter: This sentence can consistently be true and cannot consistently be false (because then both disjuncts should be false and the second cannot be), but that should not be sufficient cause for actually counting it as true, when we require groundedness and not just non-arbitrariness. That the sentence can consistently be true and cannot consistently be false is not the same as there being a grounded fact to which the sentence can correspond. At any level where that sentence itself and the Liar are undefined there is not. Hence, it should remain undefined.

the same trees also with just one evaluation, namely the one in which the end nodes are assigned values based on the evaluation of the previous level.

[21] The formalization of the sentence is $T(c_t) \vee \neg T(c_l)$ where $I(c_t) = T(c_t) \vee \neg T(c_l)$ and $I(c_l) = \neg T(c_l)$.

"It is not at all clear how the largest intrinsic fixed point fits in with the intuitive picture of truth that we get from Kripke", Gupta (1982, 37) writes, and asks rhetorically "By what sort of stage-by-stage process do we reach this fixed point?" The largest fixed point is indeed not intuitively satisfactory. We have aimed somewhat lower, and in doing so we have managed to hit the target through a stage-by-stage process.

So far so good. But we are still aiming and hitting too low. For Gupta has one more trick up his sleeve; a trick that cannot be dealt with using the approach of this chapter alone. Replace A3 and A5 with these sentences:

A3†: "A3† is true" is true
A5†: "A3† is not true" is true

Under the present theory these sentences become undefined, and as a consequence B4 does as well. But the same old reason for B4 being intuitively true seems to have undiminished power: A3† and A5† are intuitively contradictory, so the fact that they are not both true is independent of their specific truth values.

What is required to make the formal theory match the intuition is that the sentence

$$A(c_{TTA3^\dagger}) \wedge T(c_{TTA3^\dagger}) \wedge A(c_{T\neg TA3^\dagger}) \wedge T(c_{T\neg TA3^\dagger}) \rightarrow =(c_{TTA3^\dagger}, c_{T\neg TA3^\dagger})$$

(with obvious notation) is made true even though two of the conjuncts of the antecedent are undefined, the other two true, and the consequent false. And that means that compositionality must be given up. For we cannot allow that all sentences of this form with two of the conjuncts of the antecedent undefined, the other two true, and the consequent false is made true. Consider for example the sentence

$$W(c_s) \wedge T(c) \wedge W(c_s) \wedge T(c) \rightarrow B(c_s),$$

where $W(c_s)$ is the true sentence "Snow is sometimes white", $B(c_s)$ is the false sentence "Snow is always black", and c is a constant denoting the conditional itself, so that, at least initially, $T(c)$ is undefined. If we allow the conditional to be made true, then $T(c)$ would also become true, and therefore the antecedent would too. That would leave us with a true conditional with a true antecedent and a false consequent.

Having accepted that Gupta's Challenge in its first version was a good reason for modifying Kripke's theory, we have to give up on compositionality and come up with an even more extensive modification. That is the agenda for the final chapter.

That the theory of this chapter proved ultimately to be a failure does not put us back at square one. We have made some progress and gained some insights that can be used in the next attempt. First, the idea of using all available information at each stage while not using "information" that turns out to be false when the final evaluation is reached. And second, the idea about doing supervaluation that is not limited to one iteration of the truth predicate. On the other hand we will get rid of compositionality and also the rather *ad hoc* use of the strong rule for truth in some contexts and the weak rule in other contexts.

We will also have to undo a problem we have created for ourselves. We need ultimately to get rid of the reliance on classical ordinals and be able to interpret our formal theory in accordance with the notion of truth-as-potentiality. We could have done that with the basic version of Kripke's theory, but with the theory of this chapter that option seems to have been lost. Here is why: There is an alternative way to formulate the basic version using trees, instead of the one mentioned in the beginning of this section. The alternative is to allow a "full tree" to have branches containing the same sentence multiple times (i.e. remove clause 1 in the definition of "full tree" and hence also the word "otherwise" in clause 2); equip the trees (still only having finite branches) for a given sentence with only the one evaluation described above, i.e. the evaluation resulting from assigning + to all end nodes that are not labeled with an atomic sentence with an ordinary predicate or the truth predicate applied to a constant denoting a non-sentence; and count a sentence as true (false) iff some tree for it assigns $\top$ ($\perp$) to the root. This would result in the fixed point being already reached at level 1.[22] In other words, there is no need for levels at all. And we have already seen in chapter 4 how what is true or false according to a classical theory and depends on other truths and falsities in a dependency structure that has the form of a tree with finite branches can be vindicated as TAP or FAP (a sentence in a given node that is, e.g., true according to the basic version, is TAP if it is formulated by virtue of the TAP and FAP values that the sentences of the immediate successors would have if they were formulated). But where Kripke's basic theory is not essentially dependent on ordinals, it would seem that the present theory is. So it is not clear that what has been defined as true and false in this theory, formulated using classical set theory, makes sense for a non-verificationist constructivist.

[22]This was demonstrated in (Davis 1979) (although there are some inessential technical differences); see also (Hazen 1981).

Chapter 7

Supervaluation on posets

7.1 Introduction and examples

If we give up on compositionality, but insist that truths have to be grounded and that a disjunction can only be true if one of its disjuncts is, what semantics does that leave us with for disjunctions? The case of truth does not seem to leave much wiggle room, for if one of the disjuncts can be made true, then surely we have the ground for making the disjunction itself true. Thus, a disjunction is true if and only if at least one of the disjuncts is true. However, for falsity there are, I think, options.

One prominent idea about falsity – indeed the idea that the classical logician believes completely determines it – is that falsity, to speak metaphorically, should fill up the space left by truth. For reasons that have already been outlined and will be discussed in more detail in Section 7.8, I don't think that is compatible with a comprehensive theory of classes based on the idea of grounding. However, it does seem to be so compatible that falsity takes up more space than in Strong Kleene. The classical logician would have it that a sentence is false if it is not true. We could allow that a sentence is false if *it can be grounded that* it is not true.

In the Strong Kleene truth table for disjunction there is just one entry occupied by falsity and three that are taken by the value of undefined. Together they cover the area characterized by it being the case for each disjunct that it is either false or undefined. The idea to be pursued here is that we let the disjunction be false if it can be grounded that the situation for the disjuncts must be somewhere in that area. That is, if it can be established through supervaluation that each disjunct must be either false or undefined, perhaps without anything more specific than that being clear, then the disjunction is declared false. Falsity for a disjunction will be guaranteed absence of truth.

$\neg\phi$		
	$\top$	$\bot$
ϕ	$\bot$	$\top$
	$+$	${}^{\top}\!{+}$

$\phi \vee \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\top$	$\top$
ϕ	$\bot$	$\top$	$\bot$	${}^{+}\!{\bot}$
	$+$	$\top$	${}^{+}\!{\bot}$	${}^{+}\!{\bot}$

$T(c_\phi)$		
	$\top$	$\top$
ϕ	$\bot$	$\bot$
	$+$	${}^{+}\!{\bot}$

$\phi \wedge \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\bot$	${}^{+}\!{\bot}$
ϕ	$\bot$	$\bot$	$\bot$	$\bot$
	$+$	${}^{+}\!{\bot}$	$\bot$	${}^{+}\!{\bot}$

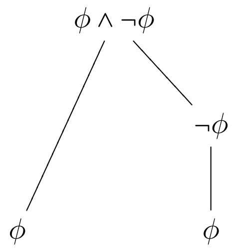
$\phi \to \psi$		ψ: $\top$	$\bot$	$+$
	$\top$	$\top$	$\bot$	${}^{+}\!{\bot}$
ϕ	$\bot$	$\top$	$\top$	$\top$
	$+$	$\top$	${}^{\top}\!{+}$	${}^{\top}\!{+}$

Figure 7.1

That way, it might be the case that a disjunction is made false with one disjunct false and the other disjunct undefined, because it was possible to establish that the second disjunct could under no circumstances end up true higher up in the hierarchy, while another disjunction is left undefined even though it also has one disjunct false and the other disjunct undefined, because the evaluations quantified over included some that made the second disjunct true. This gives us a non-compositional semantics for the disjunction as illustrated in Figure 7.1.

Figure 7.1 also contains similar truth tables for the other connectives and the truth predicate. The idea is the same: A negation is false if the negated sentence is true, false if it can be grounded that it is not true, and undefined otherwise. A conjunction is true if both conjuncts are true, false if it can be grounded that that is not the case, and undefined otherwise. Slightly more complicated and of particular interest is the conditional. A conditional is true if it can be grounded that either the consequent is true or the antecedent is not. So modus ponens is validated. And we stay true to the idea behind the classical semantics for the conditional that asserting a conditional only implies a commitment to the truth of the consequent if the antecedent is true, and otherwise no commitment is made. A conditional is false if it can be grounded that the antecedent is true and the consequent is not, and undefined otherwise. Similarly, a sentence $T(c_\phi)$, saying that ϕ is true, is true if ϕ is true, false if it can be grounded that ϕ is not true, and undefined otherwise (a non-*ad hoc* way of deciding when to use the strong rule for truth and when the weak). The existential and universal quantifiers work analogously to disjunction and conjunction, respectively.

For a very simple example of the intended use of these non-compositional truth tables, which are tentative and will be revised in Section 7.3, consider a sentence of the form $\phi \wedge \neg\phi$ and this tree for it:



As in the previous chapter, we want to consider all the three possibilities of truth, falsity and undefinedness for ϕ (assuming that ϕ is neither an atomic sentence with an ordinary predicate nor an atomic sentence with the truth predicate and a constant not denoting a sentence). We do that by starting out with an *evaluation set* for the part of the tree just consisting of the two end nodes. It contains three evaluations, one making ϕ true, one making it false and one making it undefined. This can be represented like this:

ϕ	$\top$	$\bot$	$+$

Next, we extend these evaluations to also cover $\neg\phi$. First it is considered whether it can be grounded in the available information that the negated sentence is not true, as that is the truth criterion we adopted for negations. In that case we would extend all the evaluations with truth for $\neg\phi$. But obviously it cannot: ϕ being true is one of the three possibilities. So instead we extend each of the three evaluations individually. According to the table for negation, the evaluation that makes ϕ true must be extended with the assignment of falsity to $\neg\phi$, and *vice versa* for the second evaluation. The third evaluation has ϕ undefined and here we must consider two different ways to extend it. When the negated sentence is undefined, it may be possible to ground that the negated sentence is not true (with some larger tree that has $\neg\phi$ at the root and where the kind of supervaluation exemplied below happens) or it may not. Therefore, we make two "copies" of the third evaluation and extend one with $\neg\phi$ being true and the other with it being undefined:

$\neg\phi$	$\bot$	$\top$	$\top$	$+$
ϕ	$\top$	$\bot$	$+$	

Of the four possibilities we then have for the combinations of truth values for the two conjuncts of $\phi \wedge \neg\phi$, none make both true. Thereby it is grounded that the truth criterion for this sentence is not satisfied – even though we are ignorant about the actual truth values of the constituents – and we can supervaluate it as false:

$$\begin{array}{l cc|c|c}
\phi \wedge \neg\phi & & \multicolumn{2}{c}{\bot} & \\
\neg\phi & \bot & \top & \top & + \\
\phi & \top & \bot & \multicolumn{2}{c}{+}
\end{array}$$

It is intuitively reasonable to have this sentence false no matter whether ϕ is true, false or undefined, for it is guaranteed that not both conjuncts are true. We can ground the non-truth of this sentence in facts about it logical form alone; no facts about the actual truth values of its constituents are needed. Already this extremely simple example should offer a glimpse of both how to deal with Gupta's Challenge and how to reach our ultimate goal, getting certain universally quantified sentences about all classes to come out true, but we will return to that.

The use of the non-compositional truth-tables is one of two changes I will make to the system of Chapter 6. The second has to do with getting rid of the reliance on transfinite ordinals. The formulation of the theory in terms of trees suggests a solution: place the trees on top of each other. If a sentence ϕ is made true at level 1 by a tree Tr_ϕ and another sentence ψ is then made false at level 2 by a tree Tr_ψ that has an end-node labeled ϕ, then sticking Tr_ϕ onto Tr_ψ at the ϕ node would seem to give a tree that should be capable of making ψ false in one blow. (The idea is encouraging because we have already seen in Chapter 4 that there is a connection between being tree-like in structure and being in line with the doctrine of TIC and TAP. That we will also return to.)

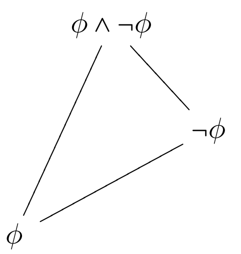
To do so, we must redefine the notion of the "full tree" for a sentence by removing the restriction that prevents the same sentence from occurring multiple times in the same branch, and that brings with it some technical problems. For, on the one hand, if we make that change *and* retain the rule that all nodes with the same label must be assigned the same value, we no longer have a theory of grounded truth, but rather a theory that, as I formulated it in the previous chapter, is searching for consistent evaluations. Consider for example this tree for the Liar:

$$\begin{array}{c}
\neg T(c_l) \\
| \\
T(c_l) \\
| \\
\neg T(c_l) \\
| \\
T(c_l) \\
| \\
\neg T(c_l)
\end{array}$$

There cannot be an evaluation of this tree that makes the end-node true, because then the node in the middle would have to be false according to the compositional rules, but true according to the rule of same label, same value. On the other hand, if we repeal this rule, then we get too many evaluations. Consider again the tree in Figure 6.1 on page 162: it would have nine evaluations instead of three, one of which would make the left-most node labeled B4 true and the right-most false, resulting in the root being false in that evaluation.

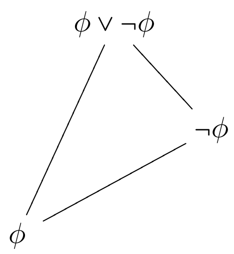
It would seem that we need to uphold the rule for "horizontal" sets of nodes with the same label, i.e. sets of nodes that are not in the same branch, but weaken it so it does not apply "vertically" to nodes in the same branch. The justification would be that upholding the rule "horizontally" just is to allow ourselves to make multiple inferences from the same assumption (e.g. inferring that A3 is true and that A5 is false from the assumption that B4 is true), while upholding the rule "vertically" is to search for a consistent evaluation and be ready to accept it merely because it is the only consistent evaluation, not because its truth values are grounded. But as stated, that is too simplistic: if there are two nodes n_1 and n_2 in the same branch plus a node n_3 not in that branch, all with the same label, then the restriction that n_1 and n_3 should have the same value and the restriction that n_2 and n_3 should have the same value would imply that n_1 and n_2 should too. A solution is to consider all the different ways the restriction could be imposed on "horizontal" sets and then quantify over them. So in the example we would check separately whether imposing the restriction on n_1 and n_3 would lead to a truth value for the root sentence and whether imposing the restriction on n_2 and n_3 would. If either does, we accept the truth value that results for that sentence.

The most elegant technical implementation of this idea is to identify the nodes that must have the same value and thereby turn the trees into partially ordered sets (posets). Thus, *in the resulting posets* there would be no need to impose any restriction about different nodes needing to have the same value. The tree for $\phi \wedge \neg\phi$ above, for example, could result in this poset, for which the above construction of an evaluation set still works:



The plan now is to first work our way through a number of examples to get a feel for the theory with as little technicality as possible, using examples where some involve a truth predicate, and then we will formulate the theory precisely but for the language of class theory introduced earlier, for that is our ultimate interest.

Beginning with the simple examples, we first satisfy ourselves that, unlike with Kripke's approach, the necessary falsehood of $\phi \wedge \neg\phi$ is not linked with the necessary truth of $\phi \vee \neg\phi$. Here is a poset for the latter:



The evaluation set for the bottom-most two nodes is as in the first example:

$$\begin{array}{ll|l|l}
\neg\phi & \bot & \top & \top \mid + \\
\phi & \top & \bot & +
\end{array}$$

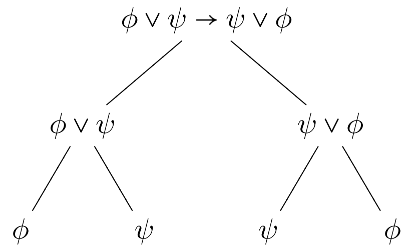
However, unlike the first example, there is no unequivocal conclusion about the truth value of $\phi \vee \neg\phi$ to be drawn; some evaluations, namely the first three, support making it true, but the last does not according to the truth table for disjunction. The evaluation set for the entire poset looks like this:

$$\begin{array}{ll|l|l}
\phi \vee \neg\phi & \top & \top & \top \mid \bot \mid + \\
\neg\phi & \bot & \top & \top \mid + \\
\phi & \top & \bot & +
\end{array}$$

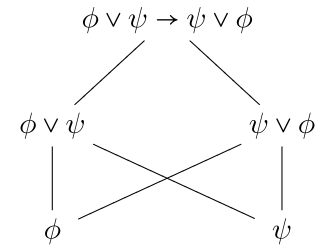
This means that $\phi \vee \neg\phi$ will only be made true if there is a larger poset for that sentence which makes it true, because it also makes either ϕ or $\neg\phi$ true.[1]

A major component of the expressibility weakness of Kripke's theory is that conditionals with undefined antecedents and consequents are themselves undefined, even when the antecedent intuitively implies the consequent. Our third example illustrates how that problem is dealt with here: a sentence of the form $\phi \vee \psi \to \psi \vee \phi$ is true in virtue of that form alone. Here is a tree for it:

[1] When we decided on the semantics for the connectives etc., we filled out more of the space left between truth and falsity by the Strong Kleene semantics with falsity than we filled out with truth. That is what accounts for this loss of duality.

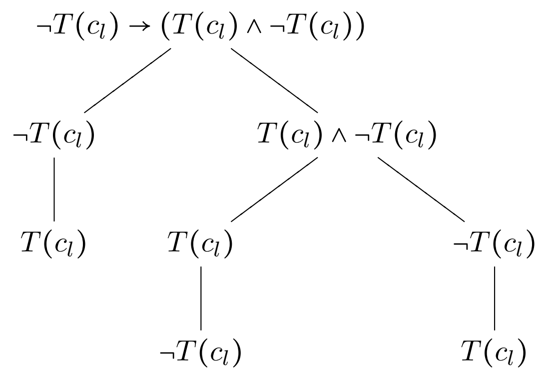


The two nodes labeled ϕ can be identified, and so can the two nodes labeled ψ, resulting in this poset:



Its evaluation set, displayed in one of the two tables in Figure 7.2, gets quite large, but all the combinations of truth values for $\phi \vee \psi$ and $\psi \vee \phi$ are such that, if the antecedent is true, then the consequent is true as well, so that the conditional is supervaluated as true.

The next example serves two purposes. First, it shows that certain semantic facts about the Liar can be expressed in the object language, a possibility that is also not present in Kripke's theory. The example is of a sentence that says that the Liar implies a contradiction: $\neg T(c_l) \rightarrow (T(c_l) \wedge \neg T(c_l))$. Second, this example illustrates the kind of situation where nodes of a tree can be identified in different ways, resulting in different posets. Here is a tree for the sentence:



Deleting the end-node labeled $\neg T(c_l)$, identifying the remaining nodes with that label and identifying the nodes labeled $T(c_l)$ results in this poset:

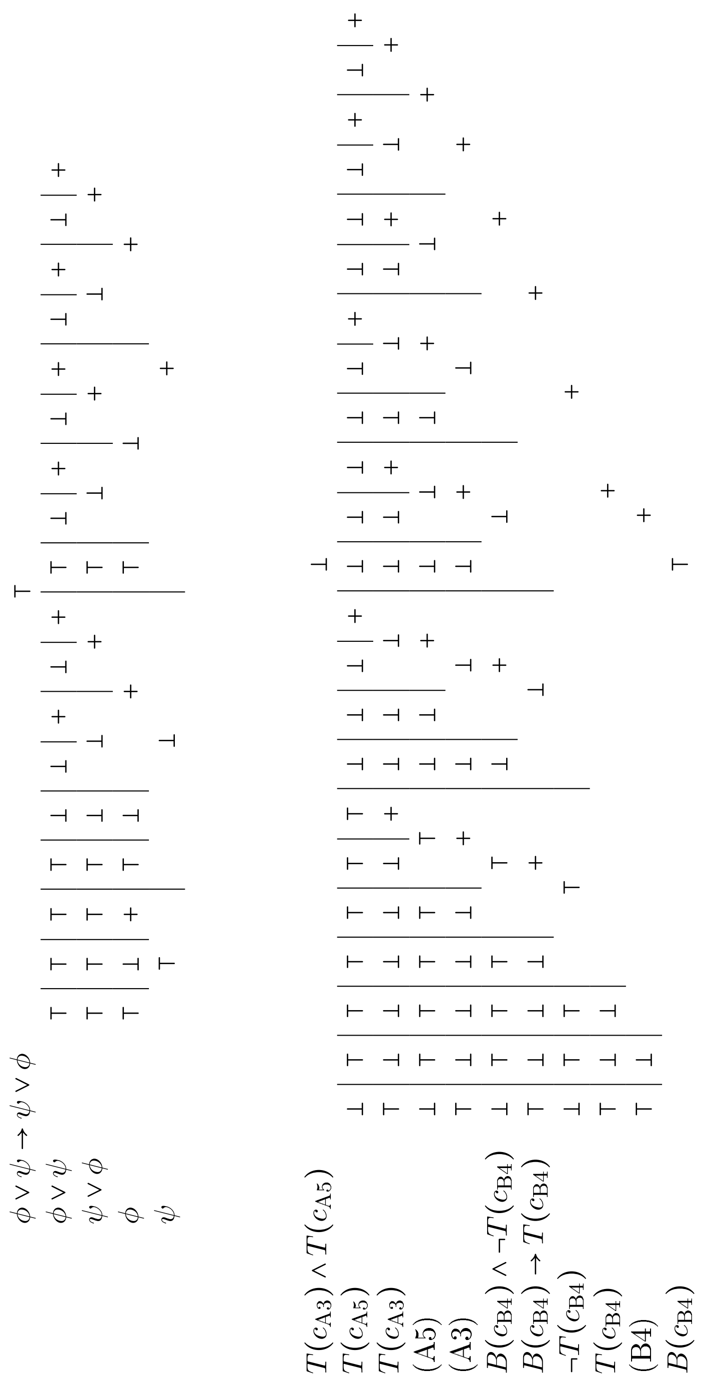
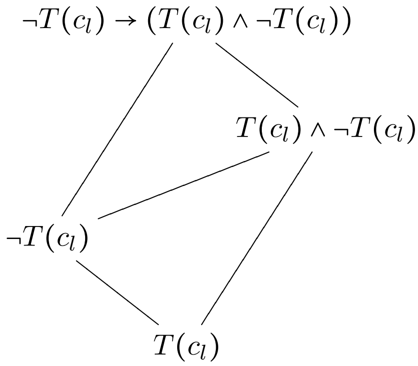


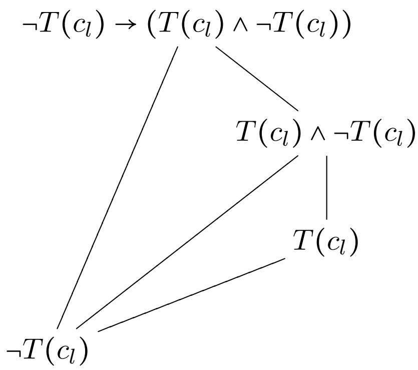
Figure 7.2



Its evaluation set is inconclusive:

$\neg T(c_l) \to (T(c_l) \wedge \neg T(c_l))$	⊤	⊥	⊥	⊥	+	⊤	+	⊤	+
$T(c_l) \wedge \neg T(c_l)$	⊥	⊥	⊥	+		⊥		+	
$\neg T(c_l)$	⊥	⊤		⊤			+		
$T(c_l)$	⊤	⊥				+			

But if, instead, we delete the two end-nodes labeled $T(c_l)$ and identify all the nodes labeled $\neg T(c_l)$, this poset results:



And with that poset, where the value of $\neg T(c_l)$ affects the value of $T(c_l)$ instead of the other way around, the sentence can be made true:

$\neg T(c_l) \to (T(c_l) \wedge \neg T(c_l))$			⊤		
$T(c_l) \wedge \neg T(c_l)$	⊤	⊥	⊥	⊥	+
$T(c_l)$	⊤	⊥	⊥	+	
$\neg T(c_l)$	⊤	⊥		+	

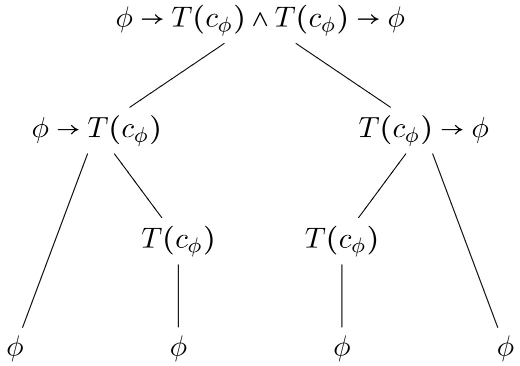
The last poset tells us that in any possible evaluation, if the antecedent is true, so is the consequent. The first poset did not contradict this. It simply did not provide us with enough information to determine it. Generalizing from this example, we have the justification for quantifying over posets and defining a sentence as true (false) if *some* poset declares it true (false), namely that posets cannot contradict each other, but some posets deliver more information than other posets (see, however, Section 7.6).

The last two examples of this section are the most interesting: Gupta's Challenge and the T-schema. Consider again the first version of Gupta's Challenge described in Chapter 6. To avoid inessential details, we just show why the sentence $T(c_{\mathrm{A3}}) \wedge T(c_{\mathrm{A5}})$ is false. It should be clear why that is practically the same as showing that $(B4)$ is true. A poset for $T(c_{\mathrm{A3}}) \wedge T(c_{\mathrm{A5}})$ looks as follows, and its evaluation set is given as the second table in Figure 7.2:

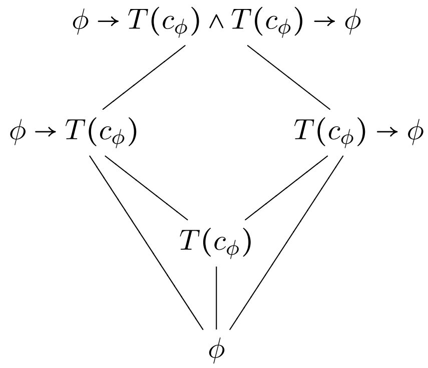
$$\begin{tikzcd}
 & & T(c_{\mathrm{A3}}) \wedge T(c_{\mathrm{A5}}) \arrow[dl, dash] \arrow[dr, dash] & & \\
 & T(c_{\mathrm{A3}}) \arrow[d, dash] & & T(c_{\mathrm{A5}}) \arrow[d, dash] & \\
 & (\mathrm{A3}) \arrow[dl, dash] \arrow[d, dash] & & (\mathrm{A5}) \arrow[d, dash] \arrow[dr, dash] & \\
\ldots & B(c_{\mathrm{B4}}) \rightarrow T(c_{\mathrm{B4}}) \arrow[dr, dash] \arrow[ddr, dash] & & B(c_{\mathrm{B4}}) \wedge \neg T(c_{\mathrm{B4}}) \arrow[dl, dash] \arrow[d, dash] & \ldots \\
 & & B(c_{\mathrm{B4}}) & \neg T(c_{\mathrm{B4}}) \arrow[dl, dash] & \\
 & & T(c_{\mathrm{B4}}) \arrow[d, dash] & & \\
 & & (B4) & &
\end{tikzcd}$$

By virtue of larger posets that have (a poset almost identical to) this poset as a sub-poset, we again get (A3) true, (A5) false and (B4) true. In the present theory, unlike in the unsuccessful theory of the previous chapter, it does not matter whether or not the bottom-most sentence is the same as the sentence that appears at the root of one of those larger posets. And that is what makes the solution stable towards the ways that the Challenge can be varied (recall the A3† example from Section 6.10 where A3† does not reappear higher up in the tree). In all variants, A3 or its analogue says, in effect, the opposite of what A5 or its analogue says about some given sentence ϕ, and that is intuitively enough to make B4 true independently of whether B4 is identical to ϕ and how the two sentences otherwise relate to each other. It is also enough in the present theory. That means that we have finally succeeded in tackling the problem at its root. In particular, all the four versions discussed in Chapter 6 get the intuitively correct truth values, as the reader can verify for him- or herself.

As explained in Section 5.2, Kripke's theory only validates a weak version of the T-schema. To repeat, in his theory $T(c)$ is true if and only if $I(c)$ is, but $\forall x(P(x) \to T(x))$ where $I(P)$ is the set of all sentences of the form $T(c_\phi) \leftrightarrow \phi$, c_ϕ designating ϕ, is undefined. In the theory of this chapter it is true. Each sentence $T(c_\phi) \leftrightarrow \phi$ has a tree of this form:



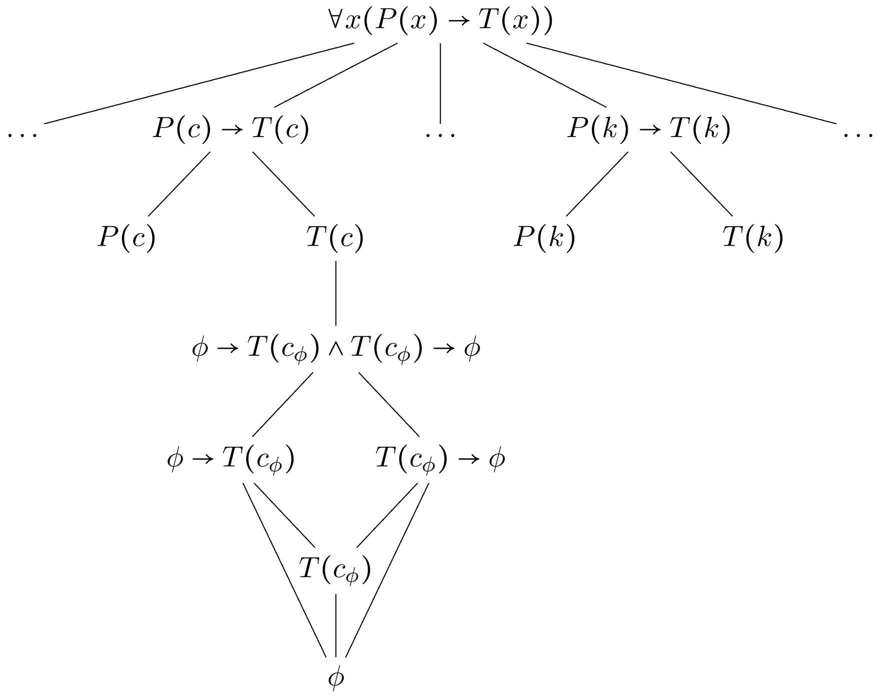
By making the proper identifications, it can be turned into a poset that looks like this:



Its evaluation set, making $T(c_\phi) \leftrightarrow \phi$ true, grounded not in the specific truth value of ϕ but in the fact that by necessity the left-hand side is true if and only if the right-hand side is true, is as follows:

$\phi \to T(c_\phi) \wedge T(c_\phi) \to \phi$	$\top$			
$\phi \to T(c_\phi)$	$\top$			
$T(c_\phi) \to \phi$	$\top$			
$T(c_\phi)$	$\top$	$\bot$	$\bot$	$+$
ϕ	$\top$	$\bot$	$+$	

The sentence $\forall x(P(x) \to T(x))$ itself has a poset which can be outlined as follows, where c represents those constants that denote a sentence of the form $T(c_\phi) \leftrightarrow \phi$, and k represents all other constants:



The evaluation set is here illustrated by pretending that there is only one of each kind of constant and by leaving out the part concerned with the sub-poset already treated:

$\forall x(P(x) \to T(x))$	$\top$
$P(c) \to T(c)$	$\top$
$P(k) \to T(k)$	$\top$
$P(c)$	$\top$
$T(c)$	$\top$
$P(k)$	$\bot$
$T(k)$	$\top \mid \bot \mid +$
$\phi \to T(c_\phi) \wedge T(c_\phi) \to \phi$	$\top$

Here we have a general law about the semantics expressed as a true sentence in the object language itself.

7.2 Considering all possibilities

As discussed in Section 6.2, we are aiming for a kind of supervaluation where "all possibilities are considered". Let us spell out how the present theory does that, partially because it is intrinsically important, partially with the aim of motivating a not yet explained detail of the theory.

Consider again the first example in this chapter, the one concerning $\phi \wedge \neg\phi$. This sentence can be made false by the creating subject at some point in time by his construction of the poset considered and its evaluation set (more detail on this in Section 7.7 below). When he does that, the idea is that no matter what else he might go on to do in terms of making sentences true and false, there will at any given time be an evaluation in the evaluation set that represents the current state. For instance, the subject may have made $\phi \wedge \neg\phi$ false at a time where both ϕ and $\neg\phi$ are undefined. In that case it is the right-most of the four evaluations (as displayed in the table) that represents the state at that time. Then later he might go on to make ϕ true and, using that, make $\neg\phi$ false. Then the right-most evaluation no longer represents the current state, but then another in the set does, namely the left-most one.[2] Alternatively he may leave ϕ undefined, but make $\neg\phi$ true by another supervaluation; that possibility is anticipated by the third of the four evaluations in the set.

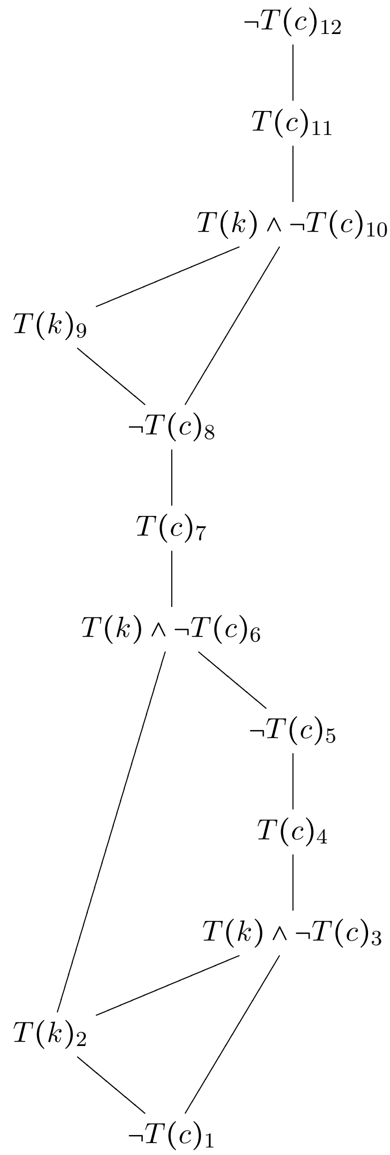
Similarly with $T(c_\phi) \leftrightarrow \phi$: ϕ may be undefined, in which case $T(c_\phi)$ is either undefined as well, or – if it has been grounded that ϕ is not true – false. Alternatively ϕ may be true, and then $T(c_\phi)$ is either true or the subject is in a position to make it true in a single step. Or ϕ can be false, in which case it similarly holds that $T(c_\phi)$ either is false or "just about to be". Those are all the possibilities for the combination of truth values for ϕ and $T(c_\phi)$ and when the subject has realized that, he can make $T(c_\phi) \leftrightarrow \phi$ true grounded in this "intensional" fact of the necessity of the sentence's truth condition being satisfied by whatever the truth values of its constituents may end up being.

To really take all possibilities into account, the theory has to be formulated carefully. Consider the sentence $T(k) \wedge \neg T(c)$ where $I(k) = \neg T(c)$ and $I(c) = T(k) \wedge \neg T(c)$. A poset for this sentence is displayed in Figure 7.3 (as there are several nodes labeled with the same sentences, they are numbered for easier reference) together with the evaluation set as it would be if values were assigned taking only the values of the immediate successors into account.

The evaluation set for nodes 1 to 5 is as it should be, but at node 6 things go wrong. It is supervaluated as false, based on the fact that there is no evaluation where the two nodes immediately below node 6 are both assigned $\top$. As a consequence, $T(c)$ is made false, $\neg T(c)$ true, $T(k)$ true and $T(k) \wedge \neg T(c)$ true. That is, an inconsistency arises.

The problem is not that the evaluation set fails to include all possibilities; what happens in lines 8 to 12 of the table is included as the left-most possibility in lines 1 to 5. The problem is that when $T(k) \wedge \neg T(c)$ is supervaluated as false, it is not the entire content of that left-most possibility that is taken into account.

[2]If the subject makes ϕ true but then happens to refrain from making $\neg\phi$ false (perhaps because he gets preoccupied with other business), then there is no evaluation in the set that fits the current situation. But then there is an evaluation that would fit, if the subject just worked his way up through the poset. The notion of an evaluation in the evaluation set representing the current situation has to be thus liberalized.



$\neg T(c)_{12}$				$\bot$			
$T(c)_{11}$				$\top$			
$T(k) \wedge \neg T(c)_{10}$				$\top$			
$T(k)_9$				$\top$			
$\neg T(c)_8$				$\top$			
$T(c)_7$				$\bot$			
$T(k) \wedge \neg T(c)_6$				$\bot$			
$\neg T(c)_5$	$\bot$	$\top$	$\top$	$\top$	$\top$	$\top$	+
$T(c)_4$	$\top$	$\bot$	$\bot$	$\bot$	$\bot$	+	
$T(k) \wedge \neg T(c)_3$	$\top$	$\bot$	$\bot$	$\bot$	+		
$T(k)_2$	$\top$	$\bot$	$\bot$	+			
$\neg T(c)_1$	$\top$	$\bot$	+				

Figure 7.3

The reason that this sentence should not be supervaluated as false is that the possibility that both $T(k)$ and $\neg T(c)$ are true should be taken into account. After all, one should be true if and only if the other is already according to the weak rule for truth! And the left-most evaluation describes exactly that situation in lines 1 and 2. But if values are assigned taking only immediate successor nodes into account, only line 5, where $\neg T(c)$ is false, is considered in the supervaluation. This example shows that we should "look further down" when the supervaluation is done. When values are to be assigned to a node labeled ϕ, we should for each constituent ψ of ϕ consider the values assigned to successors of the root labeled ψ down to nodes where the value is "determined", i.e. nodes to which all evaluations in the evaluation set assign the same value (we do not have to consider all the possibilities below a supervaluated node, for the supervaluation may eliminate some of them).[3]

The reader can easily verify that the amendment of this section does not affect any of the examples discussed in the previous section.

7.3 ...but not too many

Just as we should ensure that all possibilities are taken into account, we should strive to avoid quantifying over non-genuine possibilities, as that leads to sentences being assigned the value of undefined unnecessarily. And so far we have. I stated above that even when ϕ turns out to be undefined, it is reasonable to make $\neg\phi$ true if it can be grounded that ϕ must be either false or undefined. On that background we tentatively decided that whenever + is considered as a possible value for ϕ, both + and $\top$ must be considered possible values for $\neg\phi$, the latter because of the possibility that such grounding can take place and the former because of the possibility that it cannot. However, as a matter of fact it cannot be the case both that ϕ is undefined and that it can be grounded that ϕ is not true.

The reason is as follows. Let a poset for $\neg\phi$ be given. If it is grounded that ϕ is not true, then ϕ is assigned either + or $\perp$ in all evaluations in the evaluation set for the poset. If ϕ is a disjunction, it can be seen from Figure 7.1 that in that case the evaluation set would not contain both evaluations with + and evaluations with $\perp$. Rather, ϕ would have been supervaluated as false. So in that case the claim follows. The same holds if ϕ is a conjunction, a conditional, a quantified sentence or an atomic sentence with the truth predicate. If ϕ is

[3]Stopping at nodes that are determined ensures that the right-to-left direction of Theorem 6.7 in Chapter 6 still holds. For instance, for all sentences ϕ and ψ, if ϕ is true then so is $\phi \vee \psi$. For if ϕ is true, then there is a poset Po that makes it true, and then a poset making $\phi \vee \psi$ true can be obtained simply by taking the tree consisting of a root labeled $\phi \vee \psi$ and its two successors and sticking Po onto the node labeled ϕ.

$\neg\phi$		
	$\top$	$\perp$
ϕ	$\perp$	$\top$
	$+$	$+$

$\phi \vee \psi$		ψ: $\top$	$\perp$	$+$
	$\top$	$\top$	$\top$	$\top$
ϕ	$\perp$	$\top$	$\perp$	${}^{+}_{\perp}$
	$+$	$\top$	${}^{+}_{\perp}$	${}^{+}_{\perp}$

$T(c_\phi)$		
	$\top$	$\top$
ϕ	$\perp$	$\perp$
	$+$	$+$

$\phi \wedge \psi$		ψ: $\top$	$\perp$	$+$
	$\top$	$\top$	$\perp$	${}^{+}_{\perp}$
ϕ	$\perp$	$\perp$	$\perp$	$\perp$
	$+$	${}^{+}_{\perp}$	$\perp$	${}^{+}_{\perp}$

$\phi \to \psi$		ψ: $\top$	$\perp$	$+$
	$\top$	$\top$	$\perp$	${}^{+}_{\perp}$
ϕ	$\perp$	$\top$	$\top$	$\top$
	$+$	$\top$	${}^{\top}_{+}$	${}^{\top}_{+}$

Figure 7.4

itself a negation, then it can be seen from the figure that it is simply not possible that the evaluation set assigns $+$ to ϕ in some evaluations without also assigning $\top$ in some.

For this reason it is superfluous to consider the possibility that $\neg\phi$ is true when ϕ is undefined. Thus, in the very first example in this chapter, the one about sentences of the form $\phi \wedge \neg\phi$, we quantified over one possibility too many. Its evaluation set should instead simply look like this:

$\phi \wedge \neg\phi$		$\perp$	
$\neg\phi$	$\perp$	$\top$	$+$
ϕ	$\top$	$\perp$	$+$

Of course, that makes no difference for the truth value of this particular sentence, but it does for other sentences.

Exactly the same holds for the other type of sentence with just one constituent, atomic sentences with the truth predicate: it cannot be the case that $T(c_\phi)$ can be supervaluated as false when ϕ is undefined, and for the same reason.[4] We can thus update and simplify our truth values as shown in Figure 7.4.

This point about too many combinations being considered also applies to sentences with more than one constituent, when only one of those constituents has a "variable" truth value. Consider the case of a disjunction $\phi \vee \psi$ and a poset for it for which the node labeled ϕ immediately below the root is assigned the value $\perp$ by all evaluations in the evaluation set. (If it is assigned $\top$ by all evaluations, then $\phi \vee \psi$ is made true as well, and it cannot be the case that it is assigned $+$ by all evaluations, so $\perp$ by all evaluations is the only case to

[4] We are thus reverting to the "weak rule" for truth, as it was called in Chapter 6, after having considered using the "strong rule" in some situations in Sections 6.5-6.10 and in a different set of situations in Sections 7.1-7.2.

consider.) It cannot be the case that ψ is assigned $\perp$ or $+$ by all evaluations so that $\phi \vee \psi$ can be supervaluated as false, without it being the case that ψ is determinately false. The reason is again the same as above.

The conclusion is that when either ϕ or ψ has a determinate value in the evaluation set (i.e. all evaluations assign the same value to the node so labeled directly below the node labeled $\phi \vee \psi$), the extension of the evaluation set should be done simply using Strong Kleene, instead of splitting evaluations in further different possibilities. Further, the same holds, as the reader can verify for him- or herself, and we draw the same conclusion, in the case of conjunctions, conditionals and quantified sentences.

Another thing the reader can easily verify is that, even though the evaluation sets are to be reduced in all the examples in Section 7.1, the conclusions about the truth value of the root sentence are in all cases unaffected. An example where it does make a difference can be found below in Section 7.5.[5]

7.4 The formal theory

That concludes the motivational remarks and we can now give a formulation of the theory in classical set theory.

The language is the language of class theory introduced in Section 5.3, except that – in addition to negation, disjunction and the existential quantifier – conjunction, the conditional and the universal quantification are also taken as primitive.

As already indicated we liberalize and simplify the definition of full tree: given a sentence ξ, **the full tree for** ξ is the tree such that the root is labeled with ξ and for every node n, n has one immediate successor for each of the constituents of $l(n)$, and these successors are labeled with these constituents. **A tree for** ξ is again defined as any trimmed tree of the full tree for ξ in which all the branches are finite.

Our non-standard definition of a **poset** is as a triple $Po = (N, \prec, l)$ where N is any set, $\prec$ is a relation on N such that the transitive closure thereof, $<$, is a partial order on N satisfying the requirements that for every element of N, any linearly ordered set of predecessors of this element is finite and there is an element of N called the **root** that is a predecessor of every other element of N, and l is a function from N to $\mathcal{S}$. When $a \prec b$, we say that a is an **immediate predecessor** of b and that b is an **immediate successor** of a,

[5] A natural question to ask is why we should admit more than a single possibility as genuine. We end up with just one evaluation of all the sentences in the language, so it might seem that any other combination is not a real possibility. See footnote 11 on page 218 for an answer.

while **predecessor** (as already used in the previous sentence) and **successor** *simpliciter* is defined similarly from $<$. The terms "node", "end node" and "label" mean the same thing as in the previous chapter, as does the notation "n_ϕ". The remark about identification of isomorphic trees also carries over. A **sub-poset generated by a node** can be defined in analogy to sub-tree of a tree.

For trees, the relation $\prec$ can be defined from $<$: $n \prec m$ if $n < m$ and there is no k such that $n < k < m$.[6]

Given a poset Po and a sentence ξ, Po is **a poset for** ξ if there is a tree for ξ, Tr, and a surjective homomorphism $h : Tr \to Po$ that preserves labels, $\prec$ and $<$ and identifies identically labeled immediate successors of identified nodes. Spelled out, h has to satisfy these conditions for all nodes n, m, n' and m' of Tr:

- $l(h(n)) = l(n)$
- If $n \prec m$ then $h(n) \prec h(m)$
- If $n < m$ then $h(n) < h(m)$
- If $h(n) = h(m)$, $n \prec n'$, $m \prec m'$ and $l(n') = l(m')$ then $h(n') = h(m')$

The properties of $<$ ensure that only incomparable nodes (nodes n and m such that neither $n < m$ nor $m < n$) can be identified by a homomorphism, and identifications cannot "cross over". The effect of the last bullet point is that whenever two nodes are identified, the sub-trees they generate must also be identified.

We can now give the central definition, that of the **evaluation set** of a given poset for some sentence ξ (given a model), E. The definition is given by recursion, the base case of which is that the poset consists of just one node n with $l(n) = \xi$. In that case, E is defined as follows:

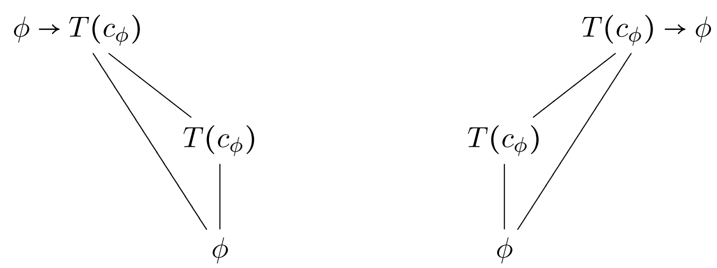
- If ξ is of the form $P(C)$ then E is the singleton of the function that has $\{n\}$ as domain and takes n to $\top$ if $I(C) \in I(P)$ and to $\bot$ if $I(C) \notin I(P)$.
- If ξ is of one the forms in E1, E2 or E3 in Section 5.3, then E is the singleton of a function that has $\{n\}$ as domain and takes n to either $\top$ or $\bot$ as indicated in those clauses.

[6] For posets, the relation $\prec$ cannot be recovered from $<$, and that is why we non-standardly take the former as primitive. If for instance b is a successor of a and c is a successor of b, we need to be able to distinguish the case where c is also an immediate successor of a from the case where it isn't.

- If ξ is of the form $t \in \{n|\phi\}$ where ϕ is a formula and t is a closed term but not a natural term, then E is the singleton of the function that has $\{n\}$ as domain and takes n to $\bot$.
- Otherwise, E is the set consisting of three functions, each with $\{n\}$ as domain and taking n to $\top$, $\bot$ and $+$, respectively.

The recursion case is a poset Po with more than one node. For each immediate successor of the root n, the evaluation set for the sub-poset generated by that immediate successor is defined. For each of those immediate successors m, let F_m be the evaluation set for the associated sub-poset. These are "combined" to form F, a set of functions on the domain consisting of the nodes of Po except n, as follows. Let M be the set of immediate successors of n. For each family $\{f_m\}_{m \in M}$ consisting of exactly one evaluation $f_m \in F_m$ for each $m \in M$, $\{f_m\}_{m \in M}$ is **in agreement** if for all nodes $m' \in Po \smallsetminus \{n\}$ there exists a truth value v such that for all $m \in M$, if $f_m(m')$ is defined, then $f_m(m') = v$. F is the set consisting of $\bigcup_{m \in M} f_m$ for each $\{f_m\}_{m \in M}$ that is in agreement. Such a set of functions is called a **root-less evaluation set**.

That was a complicated way of expressing something that is really quite simple, but that is sometimes the price of precision. Let's therefore digest it with an example before completing the definition. Let Po be the poset for $T(c_\phi) \leftrightarrow \phi$ displayed in Section 7.1. There are two immediate successors of the root and they generate two sub-posets, namely



and they each have an evaluation set with three evaluations (not four as they would according to the rules prior to Section 7.3):

$$\begin{array}{l c|c|c} \phi \to T(c_\phi) & & \top & \\ T(c_\phi) & \top & \bot & + \\ \phi & \top & \bot & + \end{array} \qquad \begin{array}{l c|c|c} T(c_\phi) \to \phi & & \top & \\ T(c_\phi) & \top & \bot & + \\ \phi & \top & \bot & + \end{array}$$

The left-most evaluation in the evaluation set for $\phi \to T(c_\phi)$ and the left-most evaluation in the evaluation set for $T(c_\phi) \to \phi$ agree on the value of all the nodes they have in common. That is, the set of those two evaluations is in agreement. So they can be combined into a function defined on all the nodes of the poset for $T(c_\phi) \leftrightarrow \phi$ except the root, namely one that sends each of those nodes to $\top$. In the same way there are two other combinations of one evaluation from each set that agree and can be combined, while there are six combinations, such as the left-most from the evaluation set for $\phi \rightarrow T(c_\phi)$ and the right-most evaluation in the evaluation set for $T(c_\phi) \rightarrow \phi$, that disagree on at least one common node and therefore cannot be combined. The three successful combinations can then be extended with a value for the root; that is specified by the rest of the definition below.

We need to squeeze in two auxiliary definitions first: First, a node n is **determined by a(n) (root-less) evaluation set** E if for all $e, e' \in E$, $e(n) = e'(n)$. Second, let $e_n^E(\phi)$ be the set of values assigned by e to successors m of n for which it holds that $l(m) = \phi$ and that there are nodes $n_1, \ldots, n_i$ that are not determined s.t. $n \prec n_1 \prec \ldots \prec n_i \prec m$. Using these, we can specify how to extend F to an evaluation set for all of Po:

- If ξ is of the form $\neg\phi$ then E is the smallest set such that the following holds for each $f \in F$:
  - If $\bot \in f_n^F(\phi)$ then $f \cup \{(n, \top)\} \in E$.
  - If $\top \in f_n^F(\phi)$ then $f \cup \{(n, \bot)\} \in E$.
  - If $+ \in f_n^F(\phi)$ then $f \cup \{(n, +)\} \in E$.
- If ξ is of the form $\phi \vee \psi$ then
  - if every function $f \in F$ is such that $\{\bot, +\} \subseteq f_n^F(\phi)$ and $\{\bot, +\} \subseteq f_n^F(\psi)$ then $E = \{f \cup (n, \bot) | f \in F\}$,
  - otherwise E is the smallest set such that the following holds for each $f \in F$:
    - If $\top \in f_n^F(\phi)$ or $\top \in f_n^F(\psi)$ then $f \cup \{(n, \top)\} \in E$.
    - If $\bot \in f_n^F(\phi)$ and $\bot \in f_n^F(\psi)$ then $f \cup \{(n, \bot)\} \in E$.
    - If ($\bot \in f_n^F(\phi)$ and $+ \in f_n^F(\psi)$) or ($+ \in f_n^F(\phi)$ and $\bot \in f_n^F(\psi)$) or ($+ \in f_n^F(\phi)$ and $+ \in f_n^F(\psi)$) then $f \cup \{(n, +)\} \in E$, and if furthermore neither n_ϕ nor n_ψ is determined, then $f \cup \{(n, \bot)\} \in E$.
- If ξ is of the form $\phi \wedge \psi$ then
  - if every function $f \in F$ is such that $\{\bot, +\} \subseteq f_n^F(\phi)$ or $\{\bot, +\} \subseteq f_n^F(\psi)$ then $E = \{f \cup (n, \bot) | f \in F\}$,
  - otherwise E is the smallest set such that the following holds for each $f \in F$:
    - If $\top \in f_n^F(\phi)$ and $\top \in f_n^F(\psi)$ then $f \cup \{(n, \top)\} \in E$.
    - If $\bot \in f_n^F(\phi)$ or $\bot \in f_n^F(\psi)$ then $f \cup \{(n, \bot)\} \in E$.

   - If ($\top \in f_n^F(\phi)$ and $+ \in f_n^F(\psi)$) or ($+ \in f_n^F(\phi)$ and $\top \in f_n^F(\psi)$) or ($+ \in f_n^F(\phi)$ and $+ \in f_n^F(\psi)$) then $f \cup \{(n, +)\} \in E$, and if furthermore neither n_ϕ nor n_ψ is determined, then $f \cup \{(n, \bot)\} \in E$.

- If ξ is of the form $\phi \to \psi$ then
  - if every function $f \in F$ is such that $\{\bot, +\} \subseteq f_n^F(\phi)$ or $\{\top\} \subseteq f_n^F(\psi)$ then $E = \{f \cup (n, \top) | f \in F\}$,
  - if every function $f \in F$ is such that $\{\top\} \subseteq f_n^F(\phi)$ and $\{\bot, +\} \subseteq f_n^F(\psi)$ then $E = \{f \cup (n, \bot) | f \in F\}$,
  - otherwise E is the smallest set such that the following holds for each $f \in F$:
    - If $\bot \in f_n^F(\phi)$ or $\top \in f_n^F(\psi)$ then $f \cup \{(n, \top)\} \in E$.
    - If $\top \in f_n^F(\phi)$ and $\bot \in f_n^F(\psi)$ then $f \cup \{(n, \bot)\} \in E$.
    - If $+ \in f_n^F(\phi)$ and ($\bot \in f_n^F(\psi)$ or $+ \in f_n^F(\psi)$) then $f \cup \{(n, +)\} \in E$, and if furthermore neither n_ϕ nor n_ψ is determined, then $f \cup \{(n, \top)\} \in E$.
    - If $\top \in f_n^F(\phi)$ and $+ \in f_n^F(\psi)$ then $f \cup \{(n, +)\} \in E$, and if furthermore neither n_ϕ nor n_ψ is determined, then $f \cup \{(n, \bot)\} \in E$.

- If ξ is of the form $\exists c\phi$ then
  - if every function $f \in F$ is such that $\{\bot, +\} \subseteq f_n^F(\phi(c/C))$ for all classes C then $E = \{f \cup (n, \bot) | f \in F\}$,
  - otherwise E is the smallest set such that the following holds for each $f \in F$:
    - If $\top \in f_n^F(\phi(c/C))$ for some C then $f \cup \{(n, \top)\} \in E$.
    - If $\bot \in f_n^F(\phi(c/C))$ for all C then $f \cup \{(n, \bot)\} \in E$.
    - If $+ \in f_n^F(\phi(c/C))$ for some C and $\bot \in f_n^F(\phi(c/C))$ for the rest then $f \cup \{(n, +)\} \in E$, and if furthermore more than one immediate successor of n is not determined, then $f \cup \{(n, \bot)\} \in E$.

- [Analogous if ξ is of the form $\exists m\phi$ or of the form $\exists q\phi$.]
- If ξ is of the form $\forall c\phi$ then
  - if every function $f \in F$ is such that $\{\bot, +\} \subseteq f_n^F(\phi(c/C))$ for some class C then $E = \{f \cup (n, \bot) | f \in F\}$,
  - otherwise E is the smallest set such that the following holds for each $f \in F$:
    - If $\top \in f_n^F(\phi(c/C))$ for all C then $f \cup \{(n, \top)\} \in E$.
    - If $\bot \in f_n^F(\phi(c/C))$ for some C then $f \cup \{(n, \bot)\} \in E$.
    - If $+ \in f_n^F(\phi(c/C))$ for some C and $\top \in f_n^F(\phi(c/C))$ for the rest then $f \cup \{(n, +)\} \in E$, and if furthermore more than one immediate successor of n is not determined, then $f \cup \{(n, \bot)\} \in E$.

- [Analogous if ξ is of the form $\forall m\phi$ or of the form $\forall q\phi$.]
- If ξ is of the form $t \in \{c|\phi\}$[7] then E is the smallest set such that the following holds for each $f \in F$:
  - If $\top \in f_n^F(\phi(c/t))$ then $f \cup \{(n, \top)\} \in E$.
  - If $\bot \in f_n^F(\phi(c/t))$ then $f \cup \{(n, \bot)\} \in E$.
  - If $+ \in f_n^F(\phi(c/t))$ then $f \cup \{(n, +)\} \in E$.
- [Analogous if ξ is of the form $t \in \{m|\phi\}$ or of the form $t \in \{q|\phi\}$.]

That concludes the recursive definition of "evaluation set", and we can make the final definition, which is the point of it all: a sentence is **true** (**false**) if there exists a poset for the sentence for which the evaluation set consists entirely of evaluations that assign the value $\top$ ($\bot$) to the root, and **undefined** otherwise.

The first thing to note about this theory is that every sentence that is true (false) according to the Chapter 5 theory is also true (false) according to this one.[8] Thus, all the positive results from Chapter 5 stand.

7.5 Expressive strength

The plan for the rest of this chapter is first, in this section, to demonstrate that the two example sentences from Section 5.7 now get the intuitively correct truth values and are thus examples of expressive strength and not weakness, then to discuss consistency, followed by an explanation of how this formal theory connects with non-verificationist constructivism and the notions of TIC and TAP, and finally to reflect on the status of the theory and defend the failure of bivalence.

The first example from Section 5.7 is the sentence $\mathbb{R}^{0+} \subseteq \mathbb{R}$, that is

$$\forall c(c \in \mathbb{R}^{0+} \to c \in \mathbb{R})$$

or spelled out even more

$$\forall c(c \in \{a|a \in \mathbb{R} \land \forall n, q(\langle n, q\rangle \in a \to q \geq -n^{-1})\} \to c \in \mathbb{R}).$$

[7] Here, and twice below, the symbol "$\in$" is used in the object language. Every other occurrence of this symbol in this list is in the meta-language.

[8] The proof is essentially the proof of theorem 2 in (Davis 1979). If a sentence is true (false) in the Chapter 5 theory, construct the "semantical tree", as defined in that paper, for it and make it into a tree in our sense of the word by adding immediate successors to nodes that already have some, in such a way that every node with successors have immediate successors for each of its constituents. This tree for the sentence is also a poset for the sentence, and it makes the sentence true (false).

A poset that makes this sentence true is outlined in Figure 7.5. The poset in question is the one that below the root has a sub-poset like the one displayed for each class.

The evaluation set for the sub-poset consisting of only the node labeled $C \in \mathbb{R}$ is, according to the base case of the recursive definition, the set

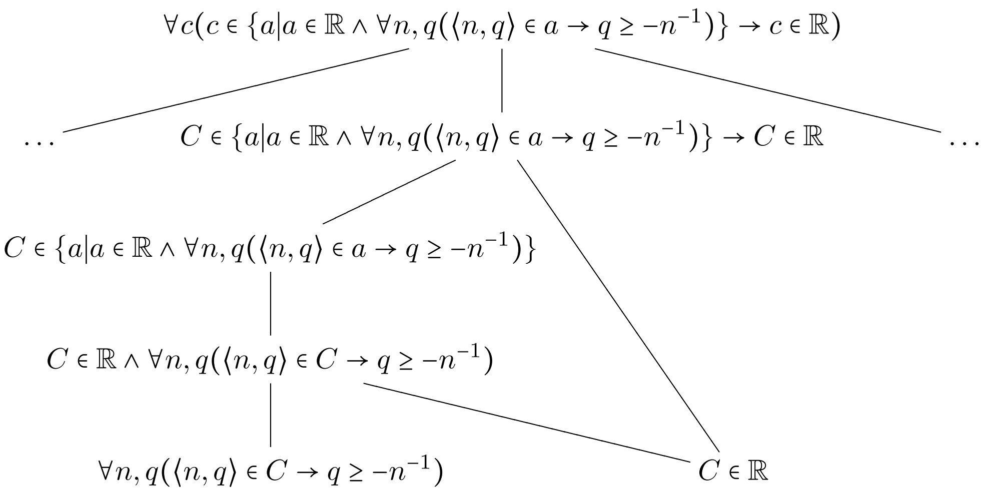
$$\{(C \in \mathbb{R}, \top), (C \in \mathbb{R}, \bot), (C \in \mathbb{R}, +)\},$$

as indicated in the last line in the table in the figure. The other end node has a similar three-element evaluation set. The root-less evaluation set for these two nodes is therefore the nine combinations of these truth value assignments, giving the penultimate line in the figure. As this root-less evaluation set includes a combination that assigns $\top$ to both nodes, it is not the first but the second of the bullet-points in the clause for conjunction that governs how to extend the root-less evaluation set to an evaluation set for the sub-poset consisting of the lower-most three nodes. As, furthermore, neither of the two end-nodes is determined, the last of the three sub-bullet points prescribes a "splitting" of the evaluations that satisfy its requirement. The result is the antepenultimate line in the table.

Going up one more node is simple; the clause for class comprehension simply tells us to extend each evaluation with the same value for that node as the value for the node below. Then, at the node with the conditional, supervaluation happens. That is, the first bullet point for conditionals applies, as it is the case that in every evaluation, either the antecedent is assigned $\bot$ or $+$ or the consequent is assigned $\top$. Thus every evaluation is extended with the value $\top$ for the node with the conditional.

As the class C is generic, every evaluation in the root-less evaluation set for the full poset minus the root assigns $\top$ to every node at the top, so the recursion clause for the universal quantifier implies that every evaluation is to be extended with $\top$ for the root. Ergo, the root sentence is true.

What we have here is a grounded truth about an indefinitely extensible concept. It is grounded because it obtains in virtue of a modal fact about all possible elements of $\mathbb{R}^{0+}$ and $\mathbb{R}$ that is independent of the truth value of the sentence itself. Notice also that even though we in this case have a *proof* of the sentence, the proof is not part of the truth maker (as it must be for any sentence about indefinitely extensible concepts with a proper truth value according to Dummett's (1991a, chapter 24) analysis). A sentence about an indefinitely extensible concept can be true or false by virtue of some poset even if that escapes our powers of proving. (More on this in Section 7.7.)



$\forall c(c \in \{a\|a \in \mathbb{R} \wedge \forall n, q(\langle n, q\rangle \in a \rightarrow q \geq -n^{-1})\} \rightarrow c \in \mathbb{R})$	$\top$
$C \in \{a\|a \in \mathbb{R} \wedge \forall n, q(\langle n, q\rangle \in a \rightarrow q \geq -n^{-1})\} \rightarrow C \in \mathbb{R}$	$\top$
$C \in \{a\|a \in \mathbb{R} \wedge \forall n, q(\langle n, q\rangle \in a \rightarrow q \geq -n^{-1})\}$	$\top \mid \bot \mid \bot \mid + \mid \bot \mid \bot \mid \bot \mid \bot \mid + \mid \bot \mid \bot \mid +$
$C \in \mathbb{R} \wedge \forall n, q(\langle n, q\rangle \in C \rightarrow q \geq -n^{-1})$	$\top \mid \bot \mid \bot \mid + \mid \bot \mid \bot \mid \bot \mid \bot \mid + \mid \bot \mid \bot \mid +$
$\forall n, q(\langle n, q\rangle \in C \rightarrow q \geq -n^{-1})$	$\top \mid \bot \mid + \mid \top \mid \bot \mid + \mid \top \mid \bot \mid +$
$C \in \mathbb{R}$	$\top \mid \bot \mid +$

Figure 7.5

The other example from Section 5.7 is the sentence

$$\forall a, b(a \in \mathbb{R} \land b \in \mathbb{R} \rightarrow a -_{\mathbb{R}} b \in \mathbb{R}).$$

This is also true, but showing that is quite a bit more complicated than in any of the previous examples, and the complexity necessitates a few preliminary remarks. The poset we will be looking at does not fit onto an A4 sheet, so the reader will find it on a large folded sheet at the end of the dissertation. As this poset also contains quantified sentences, it is again the case that the figure outlines the poset rather than actually displaying it. To save space, it has further been necessary to change notation in the case of quantified sentences: Where above, e.g. in Figure 7.5, such a sentence has been shown with three lines going down from it with three dots under two of them to indicate the infinitely many instances not shown, this figure only has one line going down from a quantified sentence. The fact that only one of infinitely many instances is shown is instead indicated by the line being broken.

To facilitate reference to the individual nodes, row and column numbers have been printed in the margin. The nodes in the top four rows, where there is only one node per row, will be referred to simply by the row number. The nodes in the rows that contain three nodes will be denoted by a name of form "r_c", where r and c are the row and column number, respectively. For nodes in the rows with more than three nodes, the format will be "r_c^n" where n, written with roman numerals, is the number of the node, counted from left to right, within the column. For instance, "12_2^i" denotes the node labeled $Q_1^B \equiv Q_3^B$. The notation will be used ambiguously for a node and its label.

Although I have not explicitly included a figure of the tree from which the poset originates, it is easy enough to discern what it looks like; just think of each of the nodes in row 20 as two nodes, one for each of its immediate predessesors. The poset is formed from this tree by not identifying any nodes at rows 1 through 19 (i.e. not any of the infinitely many instances that most of the nodes as displayed in the figure have), and by identifying row 20 nodes to the largest extent possible: a row 20 node is identified with all row 20 nodes with identical labels. In some cases this will mean that some of the six nodes that are displayed as distinct are actually identical.

The row 20 nodes require further explanation. Each instance of node 9_1^i has, in the tree, an immediate successor labeled $\langle N_1', Q_1^A\rangle \in A$, and each instance of node 19_1^i has an immediate successor labeled $\langle M^1, X_{2n}^1\rangle \in A$. The identification in node 20_1^i is meant to indicate that each instance of the former can be identified with *some* instance of the latter (namely when the existential quantifiers in nodes 14_1 and 15_1 are instantiated in the right way). Something similar holds for the other nodes in row 20. For convenience I have displayed each of these nodes with the label in both of the forms it takes in the meta-language in the immediate predessesors.

I have cheated a little bit with the notation. Except for "$^{-1}$", the superscripts in the poset are not exponents in accordance with the syntax as specified in Section 5.3, but are instead used as a supplement to subscripts.

It is not practically feasible to represent the evaluation set for this poset in the table form used so far, as we would need 3^6 columns to get through the row 20 nodes alone. Instead we will prove, in a more traditional form, that all evaluations in the evaluation set assign $\top$ to the root.

Let classes A and B be given. We need to show that sentence 3 is true in all evaluations. It is if all evaluations assign either $\bot$ or $+$ to 4 or $\top$ to 5_3, so let an evaluation be given and assume that it assigns $\top$ to 4. It follows that it assigns $\top$ to 5_1 and 5_2, and hence also to 6_1 and 6_2 and to all instances of 7_1 and 7_2 where N_1' designates an even natural number (the same of course holds when it designates an odd number, but those cases are of no interest). It again follows that there are instances of 8_1 and 8_2 that are also assigned $\top$; let us decide that it is those instances that are displayed. Going further down, it can be inferred that also 9_1^i, 9_1^{ii}, 9_2^i and 9_2^{ii}, and hence 10_1^i, all instances of 10_1^{ii}, 10_2^i and all instances of 10_2^{ii} are assigned $\top$, as are all instances of 11_1^i, 11_1^{ii}, 11_2^i and 11_2^{ii}.

Let M^1 be the same natural numeral as N_1', X_{2n}^1 the same rational numeral as Q_1^A and Y_{2n}^1 the same rational numeral as Q_1^B. From the assignment of $\top$ to 9_1^i and 9_2^i it follows that $\top$ is also assigned to 20_1^i and 20_1^{ii} and thus to 19_1^i. Let N_1 and N^1 be a natural numeral such that $I(N_1)$ and $I(N^1)$ is half of $I(M^1)$, and let Q_1 be a rational numeral such that $I(Q_1)$ equals $I(X_{2n}^1)$ minus $I(Y_{2n}^1)$. It follows that 19_1^{ii}, 18_1^i, 18_1^{ii}, 17_1, 16_1, 15_1, 14_1 and 13_1 are assigned $\top$. Ergo, 12_3^i is also assigned $\top$.

The next sub-goal is to show that 11_3^i is assigned $\top$ for the given N_1 and Q_1 and all Q_3. If $I(Q_3)$ is equal to $I(Q_1)$, then 12_3^{iii} is $\top$ so 11_3^i is too. Therefore, assume that $I(Q_3)$ is different from $I(Q_1)$, so that 12_3^{iii} is $\bot$. Let rational numerals X_{2n}^3 and Y_{2n}^3 be given. The sub-sub-goal is to show that 17_2 is $\bot$. If 18_2^{ii} is $\bot$, this follows directly. It cannot be $+$, as it is assigned either $\top$ or $\bot$ by the base case of the recursive definition of "evaluation set". Assume therefore that it is $\top$, from which it follows that $I(Q_3)$ is equal to $I(X_{2n}^3)$ minus $I(Y_{2n}^3)$. 19_2^{ii} is $\top$, so we need to show that 19_2^i is $\bot$. Let also Q_3^A be the same rational numeral as X_{2n}^3 and Q_3^B the same rational numeral as Y_{2n}^3. From $I(Q_1)$ being different from $I(Q_3)$, $I(Q_1)$ being equal to $I(Q_1^A)$ minus $I(Q_1^B)$ and $I(Q_3)$ being equal to $I(Q_3^A)$ minus $I(Q_3^B)$ it follows that either $I(Q_1^A)$ is different from $I(Q_3^A)$ or $I(Q_1^B)$ is different from $I(Q_3^B)$. The proof continues analogously in the two cases, so assume the former. In that case, 12_1^i is $\bot$. We concluded above that 11_1^i is $\top$, so it follows that 20_2^i is $\bot$.[9] Hence, also 19_2^i, 18_2^i and 17_2 are $\bot$. As both M^1, X_{2n}^3 and Y_{2n}^3 were arbitrary, and any alternative choice of N^1 would make 19_2^{ii} $\bot$, we can then infer that 16_2, 15_2, 14_2 and 13_2 are $\bot$ too, as is 12_3^{ii}. Thus, we can conclude that 11_3^i is $\top$, as we aimed for. As it is for all Q_3, also 10_3^i is $\top$, and combining this with the conclusion from the previous paragraph, we can infer that 9_3^i is $\top$.

The last sub-goal before we can reach the overall conclusion is to show that 11_3^{ii} is $\top$ for all natural numerals N_2 and rational numerals Q_2, so let such numerals be given. If 12_3^v is true, then so is 11_3^{ii}, so assume that it is false (also this node cannot be +). Let a natural numeral N^2 and rational numerals X_{2n}^2 and Y_{2n}^2 be given. Our sub-sub-goal is now to show that 17_3 is $\bot$. It is if 18_3^{ii} is $\bot$, so assume that it is $\top$ (again + is not an option). It follows that $I(N_2)$ is equal to $I(N^2)$ and that $I(Q_2)$ is equal to $I(X_{2n}^2)$ minus $I(Y_{2n}^2)$. Similarly, 17_3 is $\bot$ if 19_3^{ii} is $\bot$ so assume also that this node is $\top$, and infer that $I(M^2)$ is twice $I(N^2)$. Let N_2' be the same natural numeral as M^2, Q_2^A the same rational numeral as X_{2n}^2 and Q_2^B the same rational numeral as Y_{2n}^2. It will be demonstrated that either 20_3^i or 20_3^{ii} is $\bot$. From

$$|I(Q_2^A) - I(Q_1^A)| \leq I(N_2')^{-1} + I(N_1')^{-1}$$

and

$$|I(Q_2^B) - I(Q_1^B)| \leq I(N_2')^{-1} + I(N_1')^{-1}$$

follows

$$|I(Q_2) - I(Q_1)| \leq I(N_2)^{-1} + I(N_1)^{-1},$$

as is seen from this (recall that $I(2\cdot N^1)$ equals $I(N_1')$, $I(2\cdot N^2)$ equals $I(N_2')$, $I(N_1)$ equals $I(N^1)$ and $I(N_2)$ equals $I(N^2)$):

$$\begin{aligned}
|I(Q_2) - I(Q_1)| &= |(I(Q_2^A) - I(Q_2^B)) - (I(Q_1^A) - I(Q_1^B))| \\
&\leq |I(Q_2^A) - I(Q_1^A)| + |I(Q_2^B) - I(Q_1^B)| \\
&\leq I(N_2')^{-1} + I(N_1')^{-1} + I(N_2')^{-1} + I(N_1')^{-1} \\
&= (2\cdot I(N_2))^{-1} + (2\cdot I(N_1))^{-1} + (2\cdot I(N_2))^{-1} + (2\cdot I(N_1))^{-1} \\
&= I(N_2)^{-1} + I(N_1)^{-1}
\end{aligned}$$

By contraposition we can infer from the assumed assignment of $\bot$ to 12_3^v that either 12_1^{ii} or 12_2^{ii} is $\bot$. So as both 11_1^{ii} and 11_2^{ii} are $\top$, either 20_3^i or 20_3^{ii} is $\bot$. Thus 19_3^i, 18_3^i and 17_3 are $\bot$. From the arbitrariness of N^2, X_{2n}^2 and Y_{2n}^2 and

[9] Here and in a similar inference below the modifications in Section 7.3 are crucial: had it not been for them, 20_2^i being + would also be a possibility, and the proof would not go through.

the fact that any alternative choice of M^2 would render 19_3^{ii} $\perp$ follows then that 16_3, 15_3, 14_3 and 13_3 are as well. So 12_3^{iv} is $\perp$ and 11_3^{ii} is $\top$.

That was the sub-goal and there is now a clear path to the conclusion: The assignment of $\top$ to 11_3^{ii} has been proved for all N_2 and Q_2, so 10_3^{ii} and 9_3^{ii} are also $\top$. With the conclusion from the previous paragraph we can then infer that 8_3 is $\top$, i.e. that some instance of 8_3 is $\top$, so 7_3 is $\top$ as well. This has been proved for all N_1, so also 6_3 and hence 5_3 is $\top$. Discarding the assumption that 4 is $\top$, it can be inferred that 3 is $\top$. Also the choices of the classes A and B were arbitrary, so we can conclude that 2 and 1 are assigned $\top$. This holds for all evaluations, so $\forall a, b(a \in \mathbb{R} \wedge b \in \mathbb{R} \rightarrow a -_{\mathbb{R}} b \in \mathbb{R})$ is true.

We have thus, if nothing else, solved the two *specific* problems mentioned in Section 5.7. A sentence such as $\mathbb{R}^{0+} \subseteq \mathbb{R}$ can, even though it in a certain sense depends on both undefined sentences and itself (because a real number can be defined in such a way that one or more of its elements depend on the truth value of $\mathbb{R}^{0+} \subseteq \mathbb{R}$), be true. It is so intuitively because its truth value doesn't *really* depend on the actual truth values of all sentences of the form $C \in \mathbb{R}^{0+}$ and $C \in \mathbb{R}$, but can be grounded in the fact that any class that is an element of $\mathbb{R}^{0+}$ must by necessity also be an element of $\mathbb{R}$, and now we have a precise theory backing up that intuitive verdict.

Do we then in general have a semantics strong enough to build a decent mathematics on, that is, a semantics that even though it is trivalent is such that the undefined sentences do not have the extremely damaging consequences that they do in both intuitionism and in class theory adapted directly from Kripke's theory? I hope so. However, succeeding in establishing "the class of positive real numbers is a subclass of the class of real numbers" as true on one of the last pages of this dissertation is comparable in humorous patheticalness to the famous fact that Russell and Whitehead proved that one plus one equals two on page 379 of Principia Mathematica. That is to say that as far as a non-verificationist constructivist analysis goes, this dissertation is a mere prolegomena. Such an analysis would have to be developed in considerable detail before the adequacy of the precise theory presented in this chapter could be ascertained.

It would not come as a shock to me if it turned out that a sentence that is intuitively true or false (given non-verificationist constructivism) is undefined according to the formal theory of this chapter. My confidence, for what it is worth, in the *general approach* taken in this and the previous chapter is, however, high. I believe that if such a sentence were to be discovered, it would also be possible to blame that outcome on the formal theory quantifying over evaluations that are not genuine possibilities, in the style of Section 7.3, and thereby see a path to rectifying the formal theory.

7.6 Consistency

Unfortunately, I will also have to leave another important question open. I think that the theory is consistent but it has resisted my sustained attempts to prove it so. In this section I will explain my reason for believing that it is consistent and what I think the reaction should be if it nevertheless turns out to be inconsistent.

A heuristic argument for consistency can be build on the idea in Section 7.2. The creating subject assigns more and more truth values to sentences in a temporal process, so if he is to construct an inconsistency, one must be the first. We can divide the new truth values he assigns into two categories. First, he can assign a truth value to a sentence ξ using a poset for ξ where he in the process also assigns truth values to one or more of the constituents of ξ (by virtue of sub-posets generated by nodes that are immediate successors to the root) in such a way that the Strong Kleene prerequisites for assigning ξ the value he does, are satisfied. For instance he may make $\phi \vee \psi$ true using a poset that, in effect, also makes ϕ true. Let us call this a "Kleene assignment". Second, he may use a poset for ξ where supervaluation takes place at the root, without the Strong Kleene prerequisites being satisfied; a "supervaluation assignment".

The first inconsistency cannot be created by a sentence that has been made both true and false by Kleene assignments, for that can only happen if at least one of the constituents is already inconsistent. So at least one "half" of an inconsistency must come from a supervaluation assignment. But a supervaluation assignment only happens when the evaluation set for the rest of the poset shows, that a combination of truth values that would give the prerequisites for a Kleene assignment of the opposite truth value, is not a possibility. The idea is that during the process of making more and more sentences true and false, the subject only ever works his way up through the evaluations that are included in the evaluation set of each poset (cf. the example in Section 7.2). So if a sentence ξ is supervaluated as, say, true by some poset Po, then the process can only "satisfy" one of the evaluations in the evaluation set of Po and therefore the prerequisites for making ξ false can never be satisfied.

Of course, this presupposes that the evaluation sets genuinely include all possibilities. That is a guarantee I'm not in a position to give, and that is why the heuristic argument does not amount to a proof.

If it turns out that the theory is inconsistent, it would be because the idea of considering all possibilities when doing supervaluation is not correctly implemented. If ξ is supervaluated as, say, true, and ξ also becomes false (either by a Kleene step or by supervaluation), then that is a possibility that was ignored when ξ was made true. In that case the theory should be amended in a way that resembles the modification in Section 7.2 in such a way that the disregarded possibilities causing the inconsistency are taken into account. (And similarly, if it is discovered that, after the modifications in Section 7.3, we are still considering too many possibilities, that should also be amended.)

I am offering the formal theory of this chapter as my best suggestion on how to implement the philosophical ideas discussed. Partially because of the lack of a consistency proof, I cannot be sure that this implementation is perfect.

7.7 TIC and TAP

We have perhaps solved, and at least made considerable progress on, the problem of expressive weakness. In the process we also got rid of the reliance on classical transfinite ordinals, as the present theory is not formulated using a monotonic sequence of evaluations of the entire set of sentences, like Kripke's and the Chapter 6 theory, but instead works in such a way that the assignment of a truth value to a given sentence is done without a need to take anything but a poset for that specific sentence into account. However, as the theory has been formulated in classical set theory, we are still illegitimately piggybacking on classical mathematics, and we need to get the theory in line with the restrictions that non-verificationist constructivism places on mathematics.

My reason for first formulating the theory in classical set theory was the need to convey a precise understanding of the theory to the reader, and the contemporary dominance of classical set theory has made it a standard of precision. I have therefore employed set theory as a ladder, in the famous metaphor of Wittgenstein, and it is now time to kick it away.

Notice first that the definition of “synonymous” in Section 5.3 is entirely legitimate as it stands. The definition of synonymity between terms relies only on classical arithmetic, which was vindicated in Section 4.4, and the classical theory of rational numbers, which easily could be as well. On top of that, the definition of synonymity between closed classes relied only on finite sequences of simple syntactic transformations. Such transformations are of course also unproblematic for non-verificationist constructivism.

Considering posets and their evaluation sets, let us start out softly by again considering the first example of this chapter, the sentence $\phi \wedge \neg\phi$, and explain how it can be made false-in-content. Clearly, the existence of the three node poset for this sentence does not presuppose set theory. Constructing the three nodes, their labels and the relations of successor and immediate successor mentally is a small accomplishment, and when that is done, a sentence expressing the existence of the poset is TIC. Then the subject can construct an evaluation mentally. He can first assign the value $\top$ to the node labeled ϕ and then, following the rule for negation, assign $\perp$ to the node labeled $\neg\phi$. Similarly, he can construct the two other evaluations of the two end nodes (see Section 7.3), and when he has done that and realized that what he has constructed are all the possible evaluations of those nodes and that none of them assign $\top$ to both nodes, he can extend all the evaluations with the assignment of $\perp$ to the root. Then he has made $\phi \wedge \neg\phi$ FIC.

When a sentence requires an infinite poset in order to be made true, it cannot become true-in-content, but then truth-as-potentiality takes over, so to speak, in a way that is very similar to how it works for arithmetic, as explained in Chapter 4. Let us turn to the sentence $\mathbb{R}^{0+} \subseteq \mathbb{R}$ for an example of that; see pages 206-208. The poset described there cannot exist as a completed construction. But the subject can give himself a set of instructions that covers how any part of the poset is to be constructed. They would go like this: "Create a node labeled $\mathbb{R}^{0+} \subseteq \mathbb{R}$. Under it, place nodes with each of the instances of that sentence. For each of those nodes do ...", where the dots represent an instruction for how to create the sub-poset as shown in Figure 7.5. This description cannot be followed to completion, but as the description gives determinate instructions on how to construct any part of the poset, the poset exists as a potentially infinite structure in the sense that the class of natural numbers does (Chapter 4) and a lawless choice sequence does not (Chapter 2). For each class C, it is TAP that the sub-poset consisting of just the node labeled $C \in \mathbb{R}$ has an evaluation assigning $\top$ to its sole node, when that claim has been formulated. It is also TAP that it has evaluations with $\perp$ and with $+$. As the same holds for the other end node, it is TAP (when formulated) that the sub-poset consisting of the lower-most three nodes has an evaluation assigning $\top$ to all three nodes, and similarly for the other 11 evaluations in Figure 7.5. And of course analogous sentences are TAP for the bottom-most four nodes. So no sentence saying that there is an evaluation assigning $\top$ to the node labeled $C \in \{a | a \in \mathbb{R} \wedge \forall n, q(\langle n, q\rangle \in a \rightarrow q \geq -n^{-1})\}$ and $\perp$ or $+$ to the node labeled $C \in \mathbb{R}$ is TAP, and therefore the only sentences describing evaluations for the sub-poset consisting of the five bottom-most nodes that can be TAP when formulated are some that, among other things, assert that $\top$ is assigned to the root of that sub-poset. For any possible class C this would hold. And *that* is the truth maker for the TAP of the sentence $\mathbb{R}^{0+} \subseteq \mathbb{R}$.

Similar accounts can be given of the other examples considered above, and in this way the formal theory given in classical set theory can to a large extent be legitimized on the basis consisting of the constructivist notions of TIC and TAP. But only "to a large extent", not completely. According to the classical theory there are posets and/or evaluations of them that cannot be (finitely) described. Take for example a poset consisting of just a root labeled with a quantified sentence and immediate successors thereof labeled with each of its instances, none of which are atomic sentences. According to the classical theory this poset has an evaluation set with uncountably many ($3^{\aleph_0}$) evaluations, some of which are therefore indescribable.

We have no place for such Platonic entities, so the theory I will stand by is not identical to the one given in Section 7.4. Two modifications are necessary. The first is that a sentence is only true or false if there is a describable poset for which the evaluation set only contains evaluations that assign $\top$ to the root, and similarly for falsity. The second is that only describable evaluations are acknowledged and that the quantification in the supervaluation clauses and in the definitions of "true" and "false" are thus restricted (restricted, that is, from the perspective of the Platonist).[10]

Notice that the former restriction potentially robs some sentences of the status of being true or false, while the latter potentially assigns that honor to more sentences. I can't think of an example of a sentence that has different truth values in the classical and the constructivist versions of the theory, but it would be most interesting to know of one.

7.8 Conventionalism and failure of bivalence

That's it. The designing and modifying and amending of theories of classes that has been going on for the last three chapters is now completed – at least as far as this dissertation goes. The theory as it is now is my best recommendation for a theory of classes on which to build a non-verificationist constructivism. It is now time to take a step back and reflect on the result. What is the status of the theory? Have we now found The Correct Rules for the subject to follow?

I can think of two different senses in which a set of rules for the connectives, the quantifiers and class comprehension could be the uniquely correct one. The first is a linguistic sense: that they adequately capture what we have all meant, all along, when we used the connectives etc. The second is that the rules somehow reflect the grammar of independently given propositions. It

[10] Similarly, the interpretation function, introduced in Section 5.3, must assign to each predicate a *describable* subset of the domain, i.e. a subclass of the domain.

should be obvious that the rules are not "correct" in either of these senses. The latter would, again, require the kind of Platonistic assumption that we have long ago forsaken. And claiming that the rules are correct in the former sense would be monumentally implausible, even if the result (what sentences are declared true, false and undefined) is largely in agreement with intuition, for the ordinary language user certainly does not think in terms of trees, posets and evaluation sets when ascertaining the truth value of a sentence.

It is thus difficult to see how any set of rules for the subject could be defended as the uniquely correct one. Rather, the rules that the subject decides to follow when dealing with, e.g., "$\wedge$" constitute the meaning of that connective, and as there are different sets of rules that he could decide on, the meaning of the conjunction is not a given, but a matter of convention.

So I do not think that the Kripke-like theory of classes from Chapter 5 is *mistaken*. It is just, due to its expressive weakness, of limited usefulness. And the theory of this chapter is not the correct one, but just the most useful (or so I will believe until I am presented with a superior set of rules).

I will close this chapter by connecting this conventionalism with the subject of failure of bivalence and thus tie a knot back to the discussion of Brouwer, where I criticized his arguments against bivalence. I owe a proper and explicit justification of why my reasons for settling on a gappy semantics are better than his. Specifically – to return to the example of Russell's Class from section 5.4 – why settle on rules that have as a consequence that neither $\mathcal{R} \in \mathcal{R}$ nor $\mathcal{R} \notin \mathcal{R}$ is true? A critic may attempt to use the strong objectivist assumption that marks the difference between non-verificationist and verificationist constructivism against this class theory as follows: The idea of TAP as a concept of truth is only legitimate in so far as it is assumed that there are objective facts about which constructions are possible and which are nor, so if $\mathcal{R} \in \mathcal{R}$ cannot be made true then there is an objective fact to the effect that it is not possible to construct the truth of $\mathcal{R} \in \mathcal{R}$. Thus, we could let the subject follow the rule *make $\neg\phi$ true if it is not possible to make ϕ true* and use that to make $\mathcal{R} \notin \mathcal{R}$ true.

This is where it gets crucial that TIC is created in time. For even though it is objective what can be made TIC and FIC, we in a certain sense have to imagine the creating subject working in a universe where there are no facts one way or the other about that. Or at least that the creating subject is working in a universe where those facts are not available to him. We can "instruct" the subject to follow the rule that he can make $\mathcal{R} \in \mathcal{R}$ TIC if he has already made $\mathcal{R} \notin \mathcal{R}$ TIC and the rule that he can make it FIC if he has already made $\mathcal{R} \notin \mathcal{R}$ FIC, but we cannot instruct him to make $\mathcal{R} \in \mathcal{R}$ FIC if it is impossible to make it TIC. That is not a rule he can follow, for it is his own actions that will determine whether it is impossible to make $\mathcal{R} \in \mathcal{R}$ TIC, including his action or lack thereof to make $\mathcal{R} \in \mathcal{R}$ FIC, and thus such a fact of possibility or impossibility is not available to him. What is possibly TIC and what is not are not Platonic facts existing independently of the constructions he makes and thus they cannot be appealed to in setting up rules for him to follow.[11]

Even in the light of this explanation there may still seem to be a pressure to close the gap, so to speak, between TIC and FIC. One might retort that the creating subject is probably able to reflect on his own construction process and thus realize, through the same kind of reasoning that we employ, that there is no way for him to make $\mathcal{R} \in \mathcal{R}$ TIC. On that basis one might think he should be able to make it FIC. But if he was so allowed, then after he had made it FIC he could from that construct the TIC of the same sentence.

Let me make it clear what was *not* my point with that last sentence. I am not simply saying that the suggestion would result in a contradiction and therefore it must be wrong. That would not be to offer a proper explanation. Rather, my point is that if the subject was allowed to do that, then it would not be correct, after all, that there is no way for him to make $\mathcal{R} \in \mathcal{R}$ TIC, which was the premise.

We have to settle on one fixed set of rules for the subject to follow. When those rules result in some sentence being undefined, we may feel an urge to make it either true or false, but that can only be done by changing the rules, which might undermine the reasons we had for wanting to make the sentence either true or false.[12]

There are legitimate rules that would result in a bivalent semantics. For instance, the subject could choose to follow the rule that he at any time can make any sentence of the language of class theory TIC. However, that would completely change the meaning of each of those sentences; each one would become an empty tautology. Less drastically, bivalence could be achieved by imposing a hierarchy on the classes, and such a theory could be useful for certain purposes. But an expressively strong class theory needs, I think, to be

[11] It is because of this indeterminacy that supervaluation must happen by quantifying over more than one evaluation, cf. footnote 5 on page 201.

[12] A comparison with the sentences of Yablo's (1993) Paradox is interesting. For each natural number n, the n'th Yablo Sentence, Yn, is "For all $m > n$ the sentence Ym is not true". According to both Kripke's theory and (a truth version of) the present theory all the Yablo Sentences are undefined. But here we could, in contrast to the cases of Russell's Class and the Liar, arbitrarily decide to make, say, Y17 true based on the undefinedness of Yn for all $n > 17$. That would not be self-undermining. It would not be self-undermining as long as we stuck to rules that did indeed make all Yn for $n > 17$ undefined.

trivalent. Because of diagonalization, it would otherwise not be clear how we could express theorems about *all* real numbers.

In fact such a hierarchy has been implicitly imposed and used throughout this entire discussion of class theory. For *we* have expressed on many occasions that it is not the case that $\mathcal{R} \in \mathcal{R}$ is TIC and thus in the meta-language asserted that Russell's Class is not an element of Russell's Class. It is an artificial restriction of expressibility of the object language that allows us to do so. Recall that at the end of Section 4.4 it was explained that monotonicity was achieved by avoiding rules that made reference to such empirical matters as what happens not to have been constructed yet at a given point in time. That was an artificial restriction of expressibility that was useful for doing mathematics. It is similarly an artificial restriction of expressibility that we have all along only considered rules for the creating subject that only take his own constructions as input and do not make reference to what is true and false in *our* language. That restriction of expressibility was also useful: it made it possible to write this dissertation. Had it not been for that restriction, some sentences in the formal language we are studying would have depended on sentences in the language we are using, meaning that *we* would have been caught up in the loop, so to speak, and not in a position to take the truth values of the sentences of the object language as given, and we might not have been able to state that $\mathcal{R} \in \mathcal{R}$ is either TIC or not TIC.

Conclusion

The overall goal for this inquiry, as for so many other metaphysical inquiries, has been to avoid making excessive ontological claims, while "saving the phenomena". Let us here consider how non-verificationist constructivism scores on those two counts.

I claim to have located mathematical objects among the entities that it is *relatively* uncontroversial that one must posit anyway, that is, among parts of Being that we seem to have to stipulate as existing, already to save other phenomena than the mathematical. These ontological assumptions are that there are mental constructions and intentional directedness; that there are non-actual possibilities containing mental constructions of any finite complexity; and that actual human beings have the ability to commit to rules with a potential infinity of applications, using a sense of simplicity. One can of course have worries about these assumptions, but on balance they seem much easier to swallow than the obvious, strong ontological commitments of the Platonist, discussed in Chapter 1, and the hidden, strong ontological commitments of the intuitionist, revealed in Chapter 2.

The non-verificationist constructivist has to pay a *semantic* cost that both the Platonist and Mill (whom we discussed briefly in the beginning of Chapter 1) avoids, namely that mathematical language cannot be taken at face value but has to be interpreted with implicit modal operators. However, I find that cost negligible compared to the problems that both the Platonist and Mill face.

What does "saving the phenomena" amount to in this context? First and foremost it involves the ability to reconstruct all the mathematics that is essential to empirical science. If that cannot be done, we have a serious explanatory problem. Second, I think it should be counted as "phenomena" – i.e., as a datum that has to be accounted for, rather than something we can accept being denied by a philosophical theory – that we can quantify over, and truthfully assert many substantial things about, all real numbers and all collections, and indefinite extensibility makes that a challenge. For both of these "phenomena", I am optimistic on behalf of non-verificationist constructivism, but the results so far are too meager to allow any confident conclusions to be drawn. Beyond arithmetic, this dissertation is, as already mentioned, little more than a prolegomena to non-verificationist constructivist mathematics. Much more needs to be done.

However, some of the reasons for being optimistic become apparent if we connect back to the penultimate paragraph of Chapter 1. There we listed a number of weaknesses of intuitionism. First, non-constructive proofs were ruled out. They are not ruled out by non-verificationist constructivism, simply because "non-constructive proofs" is a misnomer and should be called "non-verificationist proofs", i.e., proofs that do not verify what the specific truth-maker is. We can accept those. Second, discontinuous functions were declared non-existing by intuitionism. They are alive and well in non-verificationist constructivism. Third, the failures of *tertium non datur* in intuitionism result, in the words of Herman Weil (1963, 54), "in an almost unbearable awkwardness". In non-verificationist constructivism there are also failures of *tertium non datur*, but they have been eliminated entirely from arithmetic and it remains to be seen whether the awkwardness in class theory might not be bearable. Fourth and last, while the intuitionist must reject many impredicative definitions, any impredicative class that one might try to define does exist as a class according to non-verificationist constructivism, i.e., as a linguistic object plus criteria of truth and falsity for sentences involving it. Some impredicative definitions breed undefined sentences. But we saw in Section 7.5 that this does not necessarily have detrimental effects on important mathematical theorems. However, again, more work needs to be done to ascertain the influence of pathological classes on the rest of mathematics.

Let me close by considering which of the philosophically important properties traditionally attributed to mathematics the queen of the sciences has according to non-verificationist constructivism. The answers rely on the considerations in Chapter 4.

Mathematical knowledge is *a priori* in the sense that mathematical knowledge is independent of perceptual experience. However, it is not independent of the inner experience of making mental constructions following rules.

It is a delicate matter whether mathematical truths are analytic. To be an analytic truth is to be true by virtue of linguistic conventions alone. The ultimate truth makers for mathematical truths are rules that are adopted by convention, but adopting conventions that apply to a potential infinity of cases relies on a non-conventional sense of simplicity, and that might be enough to categorize them as synthetic. However, the reliance on a sense of simplicity may be necessary for adopting *any* linguistic convention and thus be an essential component of such conventions. In that case, mathematical truths are analytic.

Next up is the property of being necessary. Since our sense of simplicity is presumably not necessary, and mathematics is relative to such a sense, mathematics is not necessary. It cannot be ruled out that aliens with a radically different sense of simplicity have a radically different mathematics. Nevertheless, true mathematical sentences are necessarily true, at least if sentences are individuated by meaning and not just syntax, as seems reasonable. For the meaning of a mathematical sentence is given by the accepted rules and the simplicity measure on which they are founded, and as long as they are kept fixed, no variation in the truth values of sentences is possible.

Finally, there is the question of whether mathematics is apodictic. Here the answer is clearly in the negative: as an attempt at following a rule can be different from the correct use of the rule, violations of adopted rules may well go unnoticed, so mathematical knowledge is not absolutely certain.

References

Allis, V. and T. Koetsier (1991). On some paradoxes of the infinite. *The British Journal for the Philosophy of Science 42*, 187–194.

Aristotle (1930). Physica. In *The Works of Aristotle*. Clarendon Press. Edited by W. D. Ross.

Austin, J. L. (1950). Truth. *Proceedings of the Aristotelian Society, Supplementary Volumes 24*, 111–128.

Bacon, A. (2011). A paradox for supertask decision makers. *Philosophical Studies 153*, 307–311.

Barnes, E. and R. P. Cameron (2011). Back to the future. *Philosophical Perspectives 25*, 1–26.

Becker, O. (1923). Beitrage zur phänomenologischen Begründung der Geometrie und ihrer physikalischen Anwendungen. *Jahrbuch für Philosophie und phänomenologische Forschung 6*, 385–560.

Belnap, N. (1982). Gupta's rule of revision theory of truth. *Journal of Philosophical Logic 11*, 103–116.

Benacerraf, P. (1962). Tasks, super-tasks, and the modern eleatics. *The Journal of Philosophy 59*, 765–784.

——— (1973). Mathematical truth. *The Journal of Philosophy 70*, 661–679.

Benacerraf, P. and H. Putnam (eds.) (1964). *Philosophy of Mathematics: Selected Readings*. Prentice-Hall.

Benacerraf, P. and C. Wright (1985). Skolem and the skeptic. *Proceedings of the Aristotelian Society, Supplementary Volumes 59*, 85–137.

Benardete, J. A. (1964). *Infinity: An Essay in Metaphysics*. Clarendon Press.

Bernays, P. (1935). Sur le platonisme dans les mathématiques. *L'Enseignement mathématique 1st ser. vol. 34*. Translated as "On Platonism in Mathematics" in (Benacerraf and Putnam 1964).

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica 22*, 23–36.

Bishop, E. (1967). *Foundations of Constructive Analysis*. McGraw-Hill.

Bishop, E. and D. Bridges (1985). *Constructive Analysis*. Springer.

Bolander, T. (2003). *Logical Theories for Agent Introspection.* Ph.D. thesis, Technical University of Denmark.

Boolos, G. (1971). The iterative conception of a set. *The Journal of Philosophy 68*, 215–232. Reprinted in (Benacerraf and Putnam 1964).

Bridges, D. and F. Richman (1987). *Varieties of Constructive Mathematics.* Cambridge University Press.

Brouwer, L. E. J. (1905). *Leven, Kunst en Mystiek.* Waltman. Translated (partially) as "Life, Art and Mysticism" in (Brouwer 1975).

——— (1907). *Over de Grondslagen der Wiskunde.* Ph.D. thesis, Amsterdam. Translated as "On the Foundations of Mathematics" in (Brouwer 1975).

——— (1908). De onbetrouwbaarheid der logische principes. *Tijdschrift voor wijsbegeerte 2*, 152–158. Translated as "The Unreliability of the Logical Principles" in (Brouwer 1975).

——— (1913). Intuitionism and formalism. *Bulletin of the American Mathematical Society 20*, 81–96. Reprinted in (Brouwer 1975).

——— (1918). Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten. Erster Teil: Allgemaine Mengenlehre. In *KNAW Verhandelingen* 1^e *sectie 12 no. 5.* Reprinted in (Brouwer 1975).

——— (1924a). Beweis, dass jede volle Funktion gleichmässig stetig ist. *Koninklijke Akademie van wetenschappen te Amsterdam, Proceedings of the section of sciences 27*, 189–193. Reprinted in (Brouwer 1975).

——— (1924b). Intuitionistische Zerlegung mathematische Grundbegriffe. *Jahresbericht der Deutschen Mathematiker-Vereinigung 33*, 251–256.

——— (1927). Über Definitionsbereiche von Funktionen. *Mathematische Annalen 97*, 60–75. Reprinted in (Brouwer 1975). Translated as "On the Domains of Definition of Functions" in (van Heijenoort 1967).

——— (1928). Intuitionistische Betrachtungen über den Formalismus. *Sitzungsberichte der Preussischen Akademie von Wissenschaften* pages 48–52. Reprinted in (Brouwer 1975).

——— (1947). Richtlijnen der intuitionistische wiskunde. In *KNAW Proceedings*, page 339. Translated as "Guidelines of intuitionistic mathematics" in (Brouwer 1975).

——— (1948). Consciousness, philosophy, and mathematics. In *Proceedings of the 10th International Congress of Philosophy*, pages 1235–1249. Reprinted in (Brouwer 1975).

——— (1951a). Changes in the relation between classical logic and mathematics. In *(van Stigt 1990).*

——— (1951b). Notes for a lecture. In *(van Stigt 1990).*

——— (1952). Historical background, principles and methods of intuitionism. *South African Journal of Science 49*, 139–146. Reprinted in (Brouwer 1975).

——— (1954). Points and spaces. *Canadian Journal of Mathematics 6*, 1–17. Reprinted in (Brouwer 1975).

——— (1955). The effect of intuitionism on classical algebra of logic. *Proceedings of the Royal Irish Academy 57*, 113–116. Reprinted in (Brouwer 1975).

——— (1975). *Collected Works 1*. North-Holland. Edited by A. Heyting.

——— (1981). *Brouwer's Cambridge Lectures on Intuitionism*. Cambridge University Press. Edited by D. van Dalen.

Burali-Forti, C. (1897). Una questione sui numeri transfiniti. *Rendiconti del Circolo Matematico di Palermo 11*, 154–164.

Cantor, G. (1874). Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Journal für die Reine und Angewandte Mathematik 77*, 258–262. Translated as "On an Elementary Question in the Theory of Manifolds" in (Ewald 1996).

——— (1883). *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*. Teubner. Translated as "Foundations of a general theory of manifolds" in (Ewald 1996).

——— (1891). Über eine elementare Frage der Mannigfaltigkeitslehre. *Jaresbericht der Deutschen Mathematiker-Vereinigung 1*, 75–78. Translated as "On an Elementary Question in the Theory of Manifolds" in (Ewald 1996).

——— (1962). *Gesammelte Abhandlungen*. Georg Olms Hildesheim. Edited by E. Zermelo.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Cauchy, A.-L. (1821). *Cours d'analyse de l'École royale polytechnique*. L'Imprimerie Royale. Translated as "Cauchy's Cours d'analyse: an annotated translation", Springer, 2009.

Chomsky, N. (1971). Recent contributions to the theory of innate ideas. In *The Philosophy of Language*. J. R. Searle.

Church, A. (1928). On the law of excluded middle. *Bulletin of the American Mathematical Society 34*, 75–78.

Clark, P. (1998). Dummett's argument for the indefinite extensibility of set and real number. *Grazer philosophische Studien 55*, 51–63.

Clark, P. and S. Read (1984). Hypertasks. *Synthese 61*, 387–390.

Cohen, P. (1963). The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences of the United States of America 50*, 1143–1148.

——— (1964). The independence of the continuum hypothesis, II. *Proceedings of the National Academy of Sciences of the United States of America 51*, 105–110.

Davidson, H. (1949). *A Short History of Chess*. McKay.

Davis, L. (1979). An alternative formulation of Kripke's theory of truth. *Journal of Philosophical Logic 8*, 289–296.

de Finetti, B. (1974). *Theory of Probability*. Wiley.

de Ockham, G. (1979). *Opera Theologica IV*. St. Bonaventure University. Edited by G. I. Etzkorn and F. E. Kelley.

Dedekind, R. (1872). *Stetigkeit und irrationale Zahlen*. Friedrich Vieweg und Sohn. Translated as "Continuity and Irrational Numbers" in "Essays on the Theory of Numbers", Dover Publications, 1963.

——— (1888). *Was sind und was sollen die Zahlen?* Friedrich Vieweg und Sohn. Translated as "What are and what should the numbers be?" in (Ewald 1996).

Devitt, M. (1983). Dummett's anti-realism. *The Journal of Philosophy 80*, 73–99.

Devlin, K. (1993). *The Joy of Sets*. Springer, second edition.

Dummett, M. (1978a). The philosophical basis of intuitionistic logic. In *(Dummett 1978b)*.

——— (1978b). *Truth and Other Enigmas*. Duckworth.

——— (1991a). *Frege: Philosophy of Mathematics*. Duckworth.

——— (1991b). *The Logical Basis of Metaphysics*. Duckworth.

——— (1993). *The Seas of Language*. Oxford University Press.

——— (2000). *Elements of Intuitionism*. Oxford University Press, second edition.

——— (2006). *Thought and Reality*. Oxford University Press.

Edgar, W. J. (1973). Is intuitionism *the* epistemically serious foundation for mathematics? *Philosophia Mathematica 10*, 113–133.

Epple, M. (2000). Did Brouwer's intuitionistic analysis satisfy its own epistemological standards? In V. F. Hendricks, S. A. Pedersen, and K. F. Jørgensen (eds.), *Proof Theory: Historical and Philosophical Significance*. Kluwer Academic Publishers.

Ewald, W. (ed.) (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Clarendon Press.

Feferman, S. (1979). Constructive theories of functions and classes. In M. Boffa, D. van Dalen, and K. McAloon (eds.), *Logic Colloquium 78*. North-Holland.

Fine, K. (2005). Our knowledge of mathematical objects. In T. Z. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology*, volume 1. Clarendon Press.

Fitting, M. (1986). Notes on the mathematical aspects of Kripke's theory of truth. *Notre Dame Journal of Formal Logic 27*, 75–88.

Frege, G. (1884). *Die Grundlagen der Arithmetik*. Verlag Wilhelm Koebner. Translated as "The Foundations of Arithmetic" (Second ed.) (1953). Basil Blackwell.

——— (1892). Über Sinn und Bedeutung. *Zeitschrift fur Philosophie und philosophische Kritik 100*, 25–50. Translated as "On sense and reference" in P. Geach and M. Black (eds.) (1962), Translations from the philosophical writings of Gottlob Frege. Basil Blackwell.

Gödel, K. (1940). *The Consistency of the Continuum Hypothesis*. Princeton University Press.

——— (1947). What is Cantor's continuum problem? *The American Mathematical Monthly 54*, 515–525. Revised and expanded version in (Benacerraf and Putnam 1964).

——— (1983). Russell's mathematical logic. In P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, second edition.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic 11*, 1–60.

Gupta, A. and N. Belnap (1993). *The Revision Theory of Truth*. The MIT Press.

Hansen, C. S. (2012). A Kripkean solution to paradoxes of denotation. In D. Lassiter and M. Slavkovik (eds.), *New Directions in Logic, Language, and Computation*. Springer.

——— (2014). Grounded ungroundedness. *Inquiry 57*, 216–243.

Hazen, A. (1981). Davis's formulation of Kripke's theory of truth: A correction. *Journal of Philosophical Logic 10*, 309–311.

Heck, R. (1993). Critical notice of M. Dummett "Frege: Philosophy of Mathematics". *Philosophical Quarterly 43*, 223–233.

Heine, E. (1872). Elemente der Funktionenlehre. *Journal für die reine und angewandte Mathematik 74*, 172–188.

Hellman, G. (1989). *Mathematics without Numbers*. Clarendon Press.

Herzberger, H. (1982). Notes on naive semantics. *Journal of Philosophical Logic 11*, 61–102.

Hesseling, D. E. (2003). *Gnomes in the Fog*. Birkhäuser.

Heyting, A. (1930a). Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie von Wissenschaften, Physikalisch-mathematische Klasse* pages 42–56. Translated as "The Formal Rules of Intuitionistic Logic" in (Mancosu 1998).

——— (1930b). Sur la logique intuitionniste. *Académie Royale de Belgique, Bulletin 16*, 957–963. Translated as "On Intuitionistic Logic" in (Mancosu 1998).

——— (1931). Die intuitionistische Grundlegung der Mathematik. *Erkenntnis 2*, 106–115. Translated as "The Intuitionist Foundations of Mathematics" in (Benacerraf and Putnam 1964).

——— (1935). Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie. In *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Verlag von Julius Springer.

——— (1956). *Intuitionism: An Introduction*. North-Holland.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. Penguin Books.

Husserl, E. (1964). *The Phenomenology of Internal Time-Consciousness*. Martinus Nijhoff.

Incurvati, L. (2012). How to be a minimalist about sets. *Philosophical Studies 159*, 69–87.

Kant, I. (1781). *Critik der reinen Vernunft*. Hartknoch. Translated as "Critique of Pure Reason", Cambridge University Press, 1999.

Kirkham, R. L. (1992). *Theories of Truth: A Critical Introduction*. MIT Press.

Kitcher, P. (1998). Mill, mathematics, and the naturalist tradition. In J. Skorupski (ed.), *The Cambridge Companion to Mill*. Cambridge University Press.

Kleene, S. C. (1952). *Introduction to Metamathematics*. North-Holland Publishing Company.

Kleene, S. C. and R. E. Vesley (1965). *The Foundations of Intuitionistic Mathematics*. North-Holland.

Koellner, P. (2013). The continuum hypothesis. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2013 edition.

Kolmogorov, A. N. (1932). Zur Deutung der intuitionistischen Logic. *Mathematische Zeitschrift 35*, 58–65. Translated as "On the Interpretation of Intuitionistic Logic" in (Mancosu 1998).

——— (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Verlag von Julius Springer.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy 72*, 690–716.

——— (1980). *Naming and Necessity*. Harvard University Press.

——— (1982). *Wittgenstein on Rules and Private Language*. Basil Blackwell.

Kuiper, J. J. C. (2004). *Ideas and Explorations: Brouwer's Road to Intuitionism*. Ph.D. thesis, Universiteit Utrecht.

Laraudogoitia, J. P. (1996). A beautiful supertask. *Mind 105*, 81–83.

——— (1998). Infinity machines and creation ex nihilo. *Synthese 115*, 259–265.

——— (2011). The inverse spaceship paradox. *Synthese 178*, 429–435.

Leitgeb, H. (2005). Paradox by (non-wellfounded) definition. *Analysis 65*, 275–278.

Linnebo, Ø. (2009). The individuation of the natural numbers. In O. Bueno and Ø. Linnebo (eds.), *New Waves in Philosophy of Mathematics*. Palgrave.

——— (2010). Pluralities and sets. *The Journal of Philosophy 107*, 144–164.

Maddy, P. (1983). Proper classes. *Journal of Symbolic Logic 48*, 113–139.

——— (1990). *Realism in Mathematics*. Clarendon Press.

——— (2000). A theory of sets and classes. In *Between Logic and Intuition: Essays in Honor of Charles Parsons*. Cambridge University Press.

Mancosu, P. (1998). *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*. Oxford University Press.

Martin-Löf, P. (1984). *Intuitionistic Type Theory*. Bibliopolis.

McCall, S. and D. M. Armstrong (1989). God's lottery. *Analysis 49*, 223–224.

Meinong, A. (1904). Über Gegenstandstheorie. In *Untersuchungen zur Gegenstandstheorie und Psychologie*. J. A. Barth.

Mendelson, E. (1997). *Introduction to Mathematical Logic*. Chapman & Hall, fourth edition.

Mill, J. S. (1895). *A System of Logic*. Longmans, Green, and Co.

Miller, A. and C. Wright (eds.) (2002). *Rule-Following and Meaning*. Acumen.

Ockham, W. (1969). *Predestination, God's Foreknowledge, and Future Contingents*. Appleton-Century-Crofts.

Øhrstrøm, P. (1984). Anselm, Ockham and Leibniz on divine foreknowledge and human freedom. *Erkenntnis 21*, 209–222.

Parsons, C. (1975). What is the iterative conception of set? In R. E. Butts and J. Hintikka (eds.), *Logic, Foundations of Mathematics, and Computability Theory*. Reidel. Reprinted in (Benacerraf and Putnam 1964).

——— (1997). What can we do "in principle"? In *Logic and Scientific Methods: Volume One of the Tenth International Congress of Logic, Methodology and Philosophy of Science*. Kluwer Academic Publishers.

Peano, G. (1889). *Arithmetices Principia, Nova Methodo Exposita*. Fratres Bocca. Partially translated as "The Principles of Arithmetic, Presented by a New Method" in (van Heijenoort 1967).

Peirce, C. S. (1935). *Collected Papers, volume 6*. Harvard University Press. Edited by C. Hartshorne, P. Weiss and A. W. Burks.

Placek, T. (1999). *Mathematical Intuitionism and Intersubjectivity*. Kluwer Academic Publishers.

Plato (1997). *Complete Works*. Hackett. Edited by J. M. Cooper.

Posy, C. J. (1974). Brouwer's constructivism. *Synthese 27*, 125–159.

——— (1976). Varieties of indeterminacy in the theory of general choice sequences. *Journal of Philosophical Logic 5*, 91–132.

Potter, M. (2004). *Set Theory and Its Philosophy: A Critical Introduction*. Oxford University Press.

Prawitz, D. (1974). On the idea of a general proof theory. *Synthese 27*, 63–77.

Prior, A. (1967). *Past, Present and Future*. Oxford University Press.

——— (1996). Some free thinking about time. In B. J. Copeland (ed.), *Logic and Reality*. Clarendon Press.

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science 7*, 131–193.

Quine, W. (1969). Speaking of objects. In *Ontological Relativity & Other Essays*. Columbia University Press.

Raatikainen, P. (2004). Conceptions of truth in intuitionism. *History and Philosophy of Logic 25*, 131–145.

Rosen, G. (2010). Metaphysical dependence: Grounding and reduction. In B. Hale and A. Hoffmann (eds.), *Modality*. Oxford University Press.

Russell, B. (1902). Letter to Frege. In *(van Heijenoort 1967)*, pages 124–125.

——— (1905). On denoting. *Mind, New Series 14*, 479–493.

——— (1907). On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society 4*, 29–53.

——— (1919). On propositions: What they are and how they mean. *Proceedings of the Aristotelian Society, Supplementary Volumes 2*, 1–43.

Schaffer, J. (2009). On what grounds what. In D. J. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.

Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford University Press.

——— (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press.

Shapiro, S. and C. Wright (2006). All things indefinitely extensible. In *Absolute Generality*. Oxford University Press.

Skolem, T. (1922). Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Matematikerkongressen i Helsingfors den 4–7 Juli 1922*. Akademiska Bokhandeln. Translated as "Some remarks on axiomatized set theory" in (van Heijenoort 1967).

Stanley, J. (2008). Philosophy of language. In *The Routledge Companion to Twentieth Century Philosophy*. Routledge.

Sundholm, G. (1983). Constructions, proofs and the meaning of logical constants. *Journal of Philosophical Logic 12*, 151–172.

Sundholm, G. and M. van Atten (2008). The proper explanation of intuitionistic logic: On Brouwer's demonstration of the Bar Theorem. In M. van Atten, P. Boldini, M. Bourdeau, and G. Heinzmann (eds.), *One Hundred Years of Intuitionism (1907-2007)*. Birkhäuser.

Tarski, A. (1933). *Pojęcie Prawda w Językach Nauk Dedukcyjnych*. Nakładem Towarzystwa Naukowego Warszawskiego. Translated as "The concept of truth in formalized languages" in A. Tarski (1956), Logic, semantics, metamethematics. Oxford University Press.

——— (1944). The semantic conception of truth. *Philosophy and Phenomenological Research 4*, 341–376.

Thomasson, A. L. (1999). *Fiction and Metaphysics*. Cambridge University Press.

Thomson, J. F. (1954). Tasks and super-tasks. *Analysis 15*, 1–13.

Torre, S. (2011). The open future. *Philosophy Compas 6*, 360–373.

Troelstra, A. S. (1983). Analysing choice sequences. *Journal of Philosophical Logic 12*, 197–260.

——— (1998). Concepts and axioms. *Philosophia Mathematica 6*, 195–208.

van Atten, M. (2004). *On Brouwer*. Wadsworth.

——— (2007). *Brouwer Meets Husserl*. Springer.

——— (2009). On the hypothetical judgement in the history of intuitionistic logic. In C. Glymour, W. Wang, and D. Westerstahl (eds.), *Logic,* *Methodology, and Philosophy of Science XIII: Proceedings of the 2007 International Congress in Beijing*, pages 122–136.

——— (2012). The development of intuitionistic logic. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2012 edition.

van Atten, M., D. van Dalen, and R. Tieszen (2002). Brouwer and Weyl: The phenomenology and mathematics of the intuitive continuum. *Philosophia Mathematica 10*, 203–226.

van Dalen, D. (2001). *L.E.J. Brouwer en de Grondslagen van de Wiskunde.* Epsilon.

——— (2013). *L.E.J. Brouwer: Topologist, Intuitionist, Philosopher.* Springer.

van Fraassen, B. (1966). Singular terms, truth-value gaps, and free logic. *The Journal of Philosophy 63*, 481–495.

van Heijenoort, J. (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931.* Harvard University Press.

van Stigt, W. P. (1990). *Brouwer's Intuitionism.* North-Holland.

Veblen, O. (1908). Continuous increasing functions of finite and transfinite ordinals. *Transactions of the American Mathematical Society 9*, 280–292.

von Neumann, J. (1923). Zur Einführung der transfiniten Zahlen. *Acta Litterarum ac Scientiarum Ragiae Universitatis Hungaricae Francisco-Josephinae, Sectio Scientiarum Mathematicarum 1*, 199–208.

Walicki, M. (2009). Reference, paradoxes and truth. *Synthese 171*, 195–226.

Wang, H. (1974). *From Mathematics to Philosophy.* Routledge and Kegan Paul. Chapter 6 is reprinted in (Benacerraf and Putnam 1964).

Weil, H. (1963). *Philosophy of Mathematics and Natural Science.* Atheneum.

Wells, D. (1986). *The Penguin Dictionary of Curious and Interesting Numbers.* Penguin Books.

Wen, L. (2001). Semantic paradoxes as equations. *Mathematical Intelligencer 23*, 43–48.

Wenmackers, S. and L. Horsten (2013). Fair infinite lotteries. *Synthese 190*, 37–61.

Whitehead, A. N. and B. Russell (1910). *Principia Mathematica, volume 1.* Cambridge University Press.

Williamsom, T. (2007). *The Philosophy of Philosophy.* Blackwell Publishing.

Wittgenstein, L. (1921). Logisch-philosophische Abhandlung. *Annalen der Naturphilosophie 14*, 185–262. Translated as "Tractatus Logico-Philosophicus" (1922). Kegan Paul.

——— (1953). *Philosophical Investigations.* Blackwell Publishing.

——— (1956). *Remarks on the Foundations of Mathematics.* Basil Blackwell.

Woodin, W. H. (2001a). The continuum hypothesis, part I. *Notices of the AMS 48*, 567–576.

——— (2001b). The continuum hypothesis, part II. *Notices of the AMS 48*, 681–690.

Wright, C. (1980). *Wittgenstein on the Foundations of Mathematics*. Duckworth.

——— (1982). Strict finitism. *Synthese 51*, 203–282.

Yablo, S. (1982). Grounding, dependence, and paradox. *Journal of Philosophical Logic 11*, 117–137.

——— (1985). Truth and reflection. *Journal of Philosophical Logic 14*, 297–349.

——— (1993). Paradox without self-reference. *Analysis 53*, 251–252.

——— (2000). A reply to New Zeno. *Analysis 60*, 148–151.

1		$\forall a, b(a \in \mathbb{R} \wedge b \in \mathbb{R} \to a -_{\mathbb{R}} b \in \mathbb{R})$	
2		$\forall b(A \in \mathbb{R} \wedge b \in \mathbb{R} \to A -_{\mathbb{R}} b \in \mathbb{R})$	
3		$A \in \mathbb{R} \wedge B \in \mathbb{R} \to A -_{\mathbb{R}} B \in \mathbb{R}$	
4		$A \in \mathbb{R} \wedge B \in \mathbb{R}$	
5	$A \in \mathbb{R}$	$B \in \mathbb{R}$	$A -_{\mathbb{R}} B \in \mathbb{R}$
6	$\forall n_1 \exists q_1(\langle n_1, q_1\rangle \in A \wedge \forall q_3(\langle n_1, q_3\rangle \in A \to q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in A \to \lvert q_2 - q_1\rvert \leq n_2^{-1} + n_1^{-1}))$	$\forall n_1 \exists q_1(\langle n_1, q_1\rangle \in B \wedge \forall q_3(\langle n_1, q_3\rangle \in B \to q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in B \to \lvert q_2 - q_1\rvert \leq n_2^{-1} + n_1^{-1}))$	$\forall n_1 \exists q_1(\langle n_1, q_1\rangle \in A -_{\mathbb{R}} B \wedge \forall q_3(\langle n_1, q_3\rangle \in A -_{\mathbb{R}} B \to q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in A -_{\mathbb{R}} B \to \lvert q_2 - q_1\rvert \leq n_2^{-1} + n_1^{-1}))$
7	$\exists q_1(\langle N'_1, q_1\rangle \in A \wedge \forall q_3(\langle N'_1, q_3\rangle \in A \to q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in A \to \lvert q_2 - q_1\rvert \leq n_2^{-1} + N_1'^{-1}))$	$\exists q_1(\langle N'_1, q_1\rangle \in B \wedge \forall q_3(\langle N'_1, q_3\rangle \in B \to q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in B \to \lvert q_2 - q_1\rvert \leq n_2^{-1} + N_1'^{-1}))$	$\exists q_1(\langle N_1, q_1\rangle \in A -_{\mathbb{R}} B \wedge \forall q_3(\langle N_1, q_3\rangle \in A -_{\mathbb{R}} B \to q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in A -_{\mathbb{R}} B \to \lvert q_2 - q_1\rvert \leq n_2^{-1} + N_1^{-1}))$
8	$\langle N'_1, Q^A_1\rangle \in A \wedge \forall q_3(\langle N'_1, q_3\rangle \in A \to Q^A_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in A \to \lvert q_2 - Q^A_1\rvert \leq n_2^{-1} + N_1'^{-1})$	$\langle N'_1, Q^B_1\rangle \in B \wedge \forall q_3(\langle N'_1, q_3\rangle \in B \to Q^B_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in B \to \lvert q_2 - Q^B_1\rvert \leq n_2^{-1} + N_1'^{-1})$	$\langle N_1, Q_1\rangle \in A -_{\mathbb{R}} B \wedge \forall q_3(\langle N_1, q_3\rangle \in A -_{\mathbb{R}} B \to Q_1 \equiv q_3) \wedge \forall n_2, q_2(\langle n_2, q_2\rangle \in A -_{\mathbb{R}} B \to \lvert q_2 - Q_1\rvert \leq n_2^{-1} + N_1^{-1})$
9	$\langle N'_1, Q^A_1\rangle \in A \wedge \forall q_3(\langle N'_1, q_3\rangle \in A \to Q^A_1 \equiv q_3)$ $\qquad$ $\forall n_2, q_2(\langle n_2, q_2\rangle \in A \to \lvert q_2 - Q^A_1\rvert \leq n_2^{-1} + N_1'^{-1})$	$\langle N'_1, Q^B_1\rangle \in B \wedge \forall q_3(\langle N'_1, q_3\rangle \in B \to Q^B_1 \equiv q_3)$ $\qquad$ $\forall n_2, q_2(\langle n_2, q_2\rangle \in B \to \lvert q_2 - Q^B_1\rvert \leq n_2^{-1} + N_1'^{-1})$	$\langle N_1, Q_1\rangle \in A -_{\mathbb{R}} B \wedge \forall q_3(\langle N_1, q_3\rangle \in A -_{\mathbb{R}} B \to Q_1 \equiv q_3)$ $\qquad$ $\forall n_2, q_2(\langle n_2, q_2\rangle \in A -_{\mathbb{R}} B \to \lvert q_2 - Q_1\rvert \leq n_2^{-1} + N_1^{-1})$
10	$\forall q_3(\langle N'_1, q_3\rangle \in A \to Q^A_1 \equiv q_3)$ $\qquad$ $\forall q_2(\langle N'_2, q_2\rangle \in A \to \lvert q_2 - Q^A_1\rvert \leq N_2'^{-1} + N_1'^{-1})$	$\forall q_3(\langle N'_1, q_3\rangle \in B \to Q^B_1 \equiv q_3)$ $\qquad$ $\forall q_2(\langle N'_2, q_2\rangle \in B \to \lvert q_2 - Q^B_1\rvert \leq N_2'^{-1} + N_1'^{-1})$	$\forall q_3(\langle N_1, q_3\rangle \in A -_{\mathbb{R}} B \to Q_1 \equiv q_3)$ $\qquad$ $\forall q_2(\langle N_2, q_2\rangle \in A -_{\mathbb{R}} B \to \lvert q_2 - Q_1\rvert \leq N_2^{-1} + N_1^{-1})$
11	$\langle N'_1, Q^A_3\rangle \in A \to Q^A_1 \equiv Q^A_3$ $\qquad$ $\langle N'_2, Q^A_2\rangle \in A \to \lvert Q^A_2 - Q^A_1\rvert \leq N_2'^{-1} + N_1'^{-1}$	$\langle N'_1, Q^B_3\rangle \in B \to Q^B_1 \equiv Q^B_3$ $\qquad$ $\langle N'_2, Q^B_2\rangle \in B \to \lvert Q^B_2 - Q^B_1\rvert \leq N_2'^{-1} + N_1'^{-1}$	$\langle N_1, Q_3\rangle \in A -_{\mathbb{R}} B \to Q_1 \equiv Q_3$ $\qquad$ $\langle N_2, Q_2\rangle \in A -_{\mathbb{R}} B \to \lvert Q_2 - Q_1\rvert \leq N_2^{-1} + N_1^{-1}$
12	$Q^A_1 \equiv Q^A_3$ $\qquad$ $\lvert Q^A_2 - Q^A_1\rvert \leq N_2'^{-1} + N_1'^{-1}$	$Q^B_1 \equiv Q^B_3$ $\qquad$ $\lvert Q^B_2 - Q^B_1\rvert \leq N_2'^{-1} + N_1'^{-1}$	$\langle N_1, Q_1\rangle \in A -_{\mathbb{R}} B$ $\qquad$ $\langle N_1, Q_3\rangle \in A -_{\mathbb{R}} B$ $\qquad$ $Q_1 \equiv Q_3$ $\qquad$ $\langle N_2, Q_2\rangle \in A -_{\mathbb{R}} B$ $\qquad$ $\lvert Q_2 - Q_1\rvert \leq N_2^{-1} + N_1^{-1}$
13	$\exists n, m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in A \wedge \langle m, y_{2n}\rangle \in B \wedge m \equiv 2 \cdot n \wedge \langle N_1, Q_1\rangle \equiv \langle n, x_{2n} - y_{2n}\rangle)$	$\exists n, m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in A \wedge \langle m, y_{2n}\rangle \in B \wedge m \equiv 2 \cdot n \wedge \langle N_1, Q_3\rangle \equiv \langle n, x_{2n} - y_{2n}\rangle)$	$\exists n, m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in A \wedge \langle m, y_{2n}\rangle \in B \wedge m \equiv 2 \cdot n \wedge \langle N_2, Q_2\rangle \equiv \langle n, x_{2n} - y_{2n}\rangle)$
14	$\exists m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in A \wedge \langle m, y_{2n}\rangle \in B \wedge m \equiv 2 \cdot N^1 \wedge \langle N_1, Q_1\rangle \equiv \langle N^1, x_{2n} - y_{2n}\rangle)$	$\exists m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in A \wedge \langle m, y_{2n}\rangle \in B \wedge m \equiv 2 \cdot N^1 \wedge \langle N_1, Q_3\rangle \equiv \langle N^1, x_{2n} - y_{2n}\rangle)$	$\exists m, x_{2n}, y_{2n}(\langle m, x_{2n}\rangle \in A \wedge \langle m, y_{2n}\rangle \in B \wedge m \equiv 2 \cdot N^2 \wedge \langle N_2, Q_2\rangle \equiv \langle N^2, x_{2n} - y_{2n}\rangle)$
15	$\exists x_{2n}, y_{2n}(\langle M^1, x_{2n}\rangle \in A \wedge \langle M^1, y_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1 \wedge \langle N_1, Q_1\rangle \equiv \langle N^1, x_{2n} - y_{2n}\rangle)$	$\exists x_{2n}, y_{2n}(\langle M^1, x_{2n}\rangle \in A \wedge \langle M^1, y_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1 \wedge \langle N_1, Q_3\rangle \equiv \langle N^1, x_{2n} - y_{2n}\rangle)$	$\exists x_{2n}, y_{2n}(\langle M^2, x_{2n}\rangle \in A \wedge \langle M^2, y_{2n}\rangle \in B \wedge M^2 \equiv 2 \cdot N^2 \wedge \langle N_2, Q_2\rangle \equiv \langle N^2, x_{2n} - y_{2n}\rangle)$
16	$\exists y_{2n}(\langle M^1, X^1_{2n}\rangle \in A \wedge \langle M^1, y_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1 \wedge \langle N_1, Q_1\rangle \equiv \langle N^1, X^1_{2n} - y_{2n}\rangle)$	$\exists y_{2n}(\langle M^1, X^3_{2n}\rangle \in A \wedge \langle M^1, y_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1 \wedge \langle N_1, Q_3\rangle \equiv \langle N^1, X^3_{2n} - y_{2n}\rangle)$	$\exists y_{2n}(\langle M^2, X^2_{2n}\rangle \in A \wedge \langle M^2, y_{2n}\rangle \in B \wedge M^2 \equiv 2 \cdot N^2 \wedge \langle N_2, Q_2\rangle \equiv \langle N^2, X^2_{2n} - y_{2n}\rangle)$
17	$\langle M^1, X^1_{2n}\rangle \in A \wedge \langle M^1, Y^1_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1 \wedge \langle N_1, Q_1\rangle \equiv \langle N^1, X^1_{2n} - Y^1_{2n}\rangle$	$\langle M^1, X^3_{2n}\rangle \in A \wedge \langle M^1, Y^3_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1 \wedge \langle N_1, Q_3\rangle \equiv \langle N^1, X^3_{2n} - Y^3_{2n}\rangle$	$\langle M^2, X^2_{2n}\rangle \in A \wedge \langle M^2, Y^2_{2n}\rangle \in B \wedge M^2 \equiv 2 \cdot N^2 \wedge \langle N_2, Q_2\rangle \equiv \langle N^2, X^2_{2n} - Y^2_{2n}\rangle$
18	$\langle M^1, X^1_{2n}\rangle \in A \wedge \langle M^1, Y^1_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1$ $\qquad$ $\langle N_1, Q_1\rangle \equiv \langle N^1, X^1_{2n} - Y^1_{2n}\rangle$	$\langle M^1, X^3_{2n}\rangle \in A \wedge \langle M^1, Y^3_{2n}\rangle \in B \wedge M^1 \equiv 2 \cdot N^1$ $\qquad$ $\langle N_1, Q_3\rangle \equiv \langle N^1, X^3_{2n} - Y^3_{2n}\rangle$	$\langle M^2, X^2_{2n}\rangle \in A \wedge \langle M^2, Y^2_{2n}\rangle \in B \wedge M^2 \equiv 2 \cdot N^2$ $\qquad$ $\langle N_2, Q_2\rangle \equiv \langle N^2, X^2_{2n} - Y^2_{2n}\rangle$
19	$\langle M^1, X^1_{2n}\rangle \in A \wedge \langle M^1, Y^1_{2n}\rangle \in B$ $\qquad$ $M^1 \equiv 2 \cdot N^1$	$\langle M^1, X^3_{2n}\rangle \in A \wedge \langle M^1, Y^3_{2n}\rangle \in B$ $\qquad$ $M^1 \equiv 2 \cdot N^1$	$\langle M^2, X^2_{2n}\rangle \in A \wedge \langle M^2, Y^2_{2n}\rangle \in B$ $\qquad$ $M^2 \equiv 2 \cdot N^2$
20	$\langle M^1, X^1_{2n}\rangle \in A$ $\langle N'_1, Q^A_1\rangle \in A$ $\qquad$ $\langle M^1, Y^1_{2n}\rangle \in B$ $\langle N'_1, Q^B_1\rangle \in B$	$\langle M^1, X^3_{2n}\rangle \in A$ $\langle N'_1, Q^A_3\rangle \in A$ $\qquad$ $\langle M^1, Y^3_{2n}\rangle \in B$ $\langle N'_1, Q^B_3\rangle \in B$	$\langle M^2, X^2_{2n}\rangle \in A$ $\langle N'_2, Q^A_2\rangle \in A$ $\qquad$ $\langle M^2, Y^2_{2n}\rangle \in B$ $\langle N'_2, Q^B_2\rangle \in B$
	1	2	3